

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification<sup>6</sup> :</b> C07H 21/04, C12N 15/12, 15/63, 15/85, C07K 14/435, C12Q 1/68, C12P 21/02, A61K 49/00, G01N 33/50, 33/68		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 96/24605</b> <b>(43) International Publication Date:</b> 15 August 1996 (15.08.96)
<b>(21) International Application Number:</b> PCT/US96/02331			[US/US]; 11021 Edgemere Road, Dallas, TX 75230 (US). CASTRILLON, Diego, H. [US/US]; 4859 Cedar Springs Road #247, Dallas, TX 75219 (US).
<b>(22) International Filing Date:</b> 9 February 1996 (09.02.96)			
<b>(30) Priority Data:</b> 08/386,495 10 February 1995 (10.02.95) US			<b>(74) Agent:</b> SHOLTZ, Charles, K.; Dehlinger & Associates, P.O. Box 60850, Palo Alto, CA 94306-1546 (US).
<b>(60) Parent Application or Grant</b> <b>(63) Related by Continuation</b> US Filed on 08/386,495 (CIP) 10 February 1995 (10.02.95)			<b>(81) Designated States:</b> AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
<b>(71) Applicants (for all designated States except US):</b> THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Stanford, CA 94305 (US). BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 West 7th Street, Austin, TX 78701 (US).			<b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> RYNER, Lisa, C. [US/US]; 227 Old La Honda Road, Woodside, CA 94062 (US). BAKER, Bruce, S. [US/US]; 522 Campus Drive, Stanford, CA 94305 (US). WASSERMAN, Steven, A.			

**(54) Title:** METHODS AND COMPOSITIONS FOR ALTERING SEXUAL BEHAVIOR**(57) Abstract**

Methods and compositions effective to alter the sexual or reproductive behavior of an insect are disclosed. The compositions include polynucleotides and polypeptides corresponding to the *fru* gene in *Drosophila* and its homologs in other species. Methods of identifying a compound effective to alter the reproductive behavior of an insect are also disclosed.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

**METHODS AND COMPOSITIONS FOR ALTERING SEXUAL BEHAVIOR****TECHNICAL FIELD**

This invention relates to methods and compositions for altering sexual behavior, 5 particularly sexual behavior affected by the *fruitless* gene of *Drosophila* and its homologues in other species. More specifically, the invention relates to methods and compositions employing the *fruitless* gene and its products and phenotypes, for insect pest control.

**REFERENCES**

10 Ashburner, M., DROSOPHILA: A LABORATORY MANUAL, Cold Spring Harbor Laboratory (Cold Spring Harbor, NY) (1989).

Ausubel, F.M., *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA.

Baker, B.S., *Nature* 340:521-524 (1989).

15 Bartel, P., *et al.*, *BioTechniques* 14:920-924 (1993).

Belote, J.M., and Baker, B.S., *Proc. Nat. Acad. Sci. USA* 84:8026-8030 (1987).

Boggs, R.T., *et al.*, *Cell* 50:739-747 (1987).

Brent, R., *et al.*, *Cell*, 43:729-736 (1985).

Bridges, C.B., *Science* 54:252-254 (1921).

20 Burtis, K.C., and Baker, B.S., *Cell* 56:997-1010 (1989).

Burtis, K.C., *et al.*, *EMBO J.* 10:2577-2582 (1991).

Calkins, C.O., *et al.*, *Florida Entomologist* 71:346 (1988).

Chien, C.-t., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 88:9578 (1991).

Cline, T.W., in EVOLUTIONARY MECHANISMS IN SEX DETERMINATION, (Wachtel, S.S., 25 Ed.), CRC Press, Cleveland, OH, pp. 23-36 (1988).

DiAntonio, *et al.*, *J. Neuroscience* 13:4924 (1993).

Durfee, T., *et al.*, *Genes & Development* 7:555 (1993).

Fields, S., *et al.*, *Nature* 340:245 (1989).

Frohman, M.A., *et al.*, *Proc. Natl. Acad. Sci. USA* 85:8998 (1988).

30 Frohman, M.A., in PCR PROTOCOLS (Innis, M.A., *et al.*, Eds.) Academic Press, San Diego, CA, pp. 28-38 (1990).

Giebultowicz, J.M., and Truman, J.W., *J. Comp. Neurol.* 226:87-95 (1984).

Gyuris J., *et al.*, *Cell*, 75:791-803 (1993).

Harlow, E., *et al.*, ANTIBODIES: A LABORATORY MANUAL, Cold Spring Harbor Press 35 (1988).

Harry, J.L., *et al.*, *Genetica* 87:1-6 (1992).

Hedley, M.L., and Maniatis, T., *Cell* 65:579586 (1992).

Hodgkin, J., *Nature* 344:721-728 (1990).

Hoshijima, K., *et al.*, *Science* 252:833-836 (1991).

Lawrence, P.A., and Johnston, P., *Cell* 45:505-513 (1986).

5 Lucchesi, J.C., and Manning, J.E., *Adv. Genet.* 24:371-429 (1987).

Maniatis, T., *et al.*, *Cell* 19:687-701 (1978).

Matsumoto, S.G., and Hildebrand, J.G., *Proc. Roy. Soc. Lond. B* 213:249-277 (1981).

McKeown, M., and Madigan, S.J., *Curr. Opin. Cell Biol.* 4:948-954 (1992).

Miller, *et al.*, *Science* 237:775 (1987).

10 Moses, *et al.*, *Nature* 340:531 (1989).

Mullis, K.B., *et al.*, U.S. Patent No. 4,683,195, issued 28 July 1987.

Mullis, K.B., U.S. Patent No. 4,683,202, issued July 28, 1987.

Nagoshi, R.N., *et al.*, *Cell* 53:229-236 (1988).

Possidente, D.R., and Murphey, R.K., *Devel. Biol.* 132:448-457 (1989).

15 Ryner, L.C., and Baker, B.S., *Genes Devel.* 5:2071-2085 (1991).

Sambrook, J., *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, Second Edition, Cold Spring Harbor Laboratory (Cold Spring Harbor, NY) (1989).

Sanger, F., *et al.*, *PNAS* 74:5463-67 (1977).

Schneider, I., *J. Embryol. Exp. Morphol.* 27:353 (1972).

20 Siegel, R.W., *et al.*, *Behav. Genet.* 14:383-410 (1984).

Slee, R., and Bownes, M., *Quart. Rev. Biol.* 65:175-204 (1990).

Smith, D.B., *et al.*, *Gene* 67:31 (1988).

Speith, H.T., *Ann. Rev. Entomol.* 19:385 (1974).

Stocker, R.F., and Gendre, N., *Devel. Biol.* 127:12-24 (1988).

25 Taylor, B.J., *Genetics* 132:179-191 (1992).

Taylor, B.J., *J. Neurogenet.* 5:173-192 (1989a).

Taylor, B.J., *J. Neurogenet.* 5:193-213 (1989b).

Taylor, B.J., *J. Neurogenet.* 8:251 (abstract) (1993).

Taylor, B.J., and Truman, J.W., *Development* 114:625-642 (1992).

30 Technau, G.M., *J. Neurogenet.* 1:113-126 (1984).

Thorn, R.S., and Truman, J.W., *J. Comp. Neurol.* 184:489-503 (1989).

Wolfner, M.F., *Trends Genet.* 4:333-337 (1988).

Wong, T.T.Y., *et al.*, *J. Econ. Entomol.* 85:1671 (1992).

Yang, X., *et al.*, *Science* 257:680 (1992).

35 Zollman, S., *et al.*, *PNAS* 91:10717 (1994).

**BACKGROUND OF THE INVENTION**

Insect pests account for massive economic losses in agriculture and pose health risks to millions of individuals. Traditional strategies for control of insects include chemical and biological approaches. Chemical approaches typically employ any of a variety of pesticides, 5 each with varying degrees of toxicity to non-insect animals. Biological approaches typically utilize naturally-occurring organisms pathogenic to insects or the development of crops that are more resistant to insects.

With an increased understanding of the mechanisms underlying insect behavior, and how these mechanisms relate to similar processes in other animals, it has become possible to 10 develop hybrid approaches to insect pest control. One type of hybrid approach involves the release of sterile individuals into the environment. Such sterile release programs have been successful at significantly reducing insect populations (see, for example, Wong, *et al.*, and Calkins, *et al.*).

**15 SUMMARY OF THE INVENTION**

In one aspect, the invention includes a substantially isolated FRU polynucleotide. In one embodiment, the polynucleotide is highly homologous to a polynucleotide derived from an insect belonging to the phylum Arthropoda. In another embodiment, the polynucleotide is highly homologous to a polynucleotide derived from an insect belonging the order Diptera. 20 In a related embodiment, the polynucleotide is highly homologous to a polynucleotide derived from an insect selected from the group consisting of medfly, fruit fly (e.g., *Drosophila*), tse-tse fly, sand fly, blowfly, flesh fly, face fly, housefly, screw worm-fly, stable fly, mosquito, and northern cattle grub. In other embodiments, the polynucleotide contains the sequence represented as SEQ ID NO:9 or SEQ ID NO:14. In related embodiments, the polynucleotide 25 encodes a FRU polypeptide having the sequence represented as SEQ ID NO:10 or SEQ ID NO:15.

In a related aspect, the invention includes a substantially isolated FRU polypeptide. In one embodiment, the polypeptide is highly homologous to a polypeptide derived from an insect belonging to the phylum Arthropoda. In another embodiment, the polypeptide is 30 highly homologous to a polypeptide derived from an insect belonging the order Diptera. In a related embodiment, the polypeptide is highly homologous to a polypeptide derived from an insect selected from the group consisting of medfly, fruit fly (e.g., *Drosophila*), tse-tse fly, sand fly, blowfly, flesh fly, face fly, housefly, screw worm-fly, stable fly, mosquito, and northern cattle grub. In other embodiments, the polypeptide contains the sequence 35 represented as SEQ ID NO:10 or SEQ ID NO:15.

In another aspect, the present invention includes an expression system and a method of producing a FRU polypeptide. The method includes introducing into a suitable host a recombinant expression system containing a FRU polynucleotide having an open reading frame (ORF), where the ORF has a polynucleotide sequence which encodes a FRU polypeptide, and wherein the ORF is operably linked to a control sequence which is compatible with a desired host. The vector is designed to express the FRU polypeptide in the selected host when the host is cultured under conditions resulting in the expression of the ORF sequence. A number of expression systems can be employed, including insect expression vectors such as baclovirus vectors, a lambda gt11 expression system with an 10 *Escherichia coli* host, and other yeast, mammalian cell and bacterial expression vectors.

The expressed FRU protein may be isolated by a variety of known methods, depending on the expression system employed. For example, a beta-gal-FRU fusion protein may be isolated by standard affinity methods employing an anti-beta-gal antibody. The FRU polynucleotide sequence may be modified so as to result in the expression of a mutant 15 polypeptide (fru) which may give rise to a dominant mutant phenotype when expressed in an insect host. Mutants generated as described above may be used to generate transgenic insects with altered sexual or reproductive behavior (e.g., sterile insects useful for insect control).

In yet another aspect, the present invention includes both polyclonal and monoclonal antibodies directed against FRU epitopes, or against epitopes encoded by a portion of the 20 sequence presented as SEQ ID NO:9 or SEQ ID NO:14. Such antibodies may be used in co-immunoprecipitation methods to identify proteins and/or nucleic acids that interact with the FRU protein and are involved in controlling sexual behavior. The antibodies may also be used to identify target genes whose transcription is regulated by FRU polypeptide. Once 25 identified, the regulatory regions of the genes may be incorporated into reporter constructs and used to screen for compounds which inhibit the interaction of the FRU polypeptide with the regulatory sequences. Such compounds may be useful as insect control agents.

Also included in the invention is a method of identifying a compound effective to alter the reproductive behavior of a target insect. The method includes (i) treating an insect cell, obtained from a target insect and carrying an expression vector containing FRU regulatory 30 sequences operably linked to a reporter gene, with a test compound, (ii) evaluating the level of expression of the reporter gene in the treated cell, and (iii) identifying the compound as effective if the compound significantly decreases the expression of the reporter gene in the treated cell relative to the expression of the reporter gene in untreated cells carrying the expression vector.

In one embodiment, the target insect belongs to the phylum Arthropoda. In another embodiment, the target insect belongs to the order Diptera. In a related embodiment, the target insect is selected from the group consisting of medfly, fruit fly (e.g., *Drosophila*), tse-tse fly, sand fly, blowfly, flesh fly, face fly, housefly, screw worm-fly, stable fly, mosquito, 5 and northern cattle grub. In another embodiment, the insect is a *Drosophila* species, and the cells are selected from the group consisting of Schneider's Line 2 and *Drosophila* Kc cells. In one embodiment, the reporter gene encodes a protein selected from the group consisting of chloramphenicol acetyl-transferase (CAT),  $\beta$ -galactosidase ( $\beta$ -gal) and luciferase.

10 These and other objects and features of the invention will become more fully apparent when the following detailed description is read in conjunction with the accompanying drawings.

#### BRIEF DESCRIPTION OF THE FIGURES

15 Figure 1 presents a schematic of a possible sexual differentiation hierarchy in *Drosophila*. Figures 2A and 2B show images of a Southern (*Drosophila* DNA) blot probed with a  $3 \times$  *dsx* repeats probe. The blot in Fig. 2A was washed at 47°C, while the blot in Fig. 2B was washed at 51°C.

20 Figures 3A and 3B show images of a Southern blot containing DNA from a set of *Drosophila* genomic clones probed with a  $3 \times$  *dsx* repeats probe (Fig. 3A) or with a second probe containing 5 *dsx* repeats (Fig. 3B).

Figure 4 presents the partial nucleotide sequence of a ~600 bp *EcoRI* DNA fragment isolated from clone  $\lambda$ Ch4A-11.

25 Figures 5A and 5B present images of Northern (sex-specific *Drosophila* poly(A)+ RNA) blots probed with the ~600 kb *EcoRI* DNA fragment shown in Fig. 4, and washed at 40°C (Fig. 5A) or 65°C (Fig. 5B).

Figure 6A shows a schematic of the ~600 bp *EcoRI* genomic DNA fragment shown in Fig. 4, indicating the positions of primers fru-1 (1) and fru-2 (2).

30 Figure 6B shows a schematic of a male-specific 3' RACE product, indicating the positions of primers fru-2 (2) and fru-5-rev.

Figure 6C shows a schematic of a female-specific 3' RACE product, indicating the positions of primers fru-2 (2) and fru-4-rev.

Figure 7A shows a schematic of the DNA fragments (f10A, f9A, f3A, f2A, f1D, f1H, f4B, f5C and f7A) isolated as part of a genomic walk spanning the *fru* locus at position 91B

of the third chromosome, as well as a schematic of the location of the HX1 cosmid, relative to the map of the *91B* region shown in Fig. 7B.

Figure 7B shows a schematic of the *91B* region of chromosome 3, indicating the positions of known *fru* lesions (mutants *fru-2*, *fru-4*, *fru-3* and *fru-1*).

5 Figure 7C shows a schematic of two *fru* deficiencies, Df(3R)P14 and Df(3R)ChaM5, relative to the map of the *91B* region shown in Fig. 7B.

Figures 7D, 7E, 7F, 7G and 7H show schematic diagrams of the location of sequences comprising five *fru* cDNA transcripts relative to the map of the *91B* region shown in Fig. 7B. Exons are indicated as boxes and introns as lines.

10 Figure 8 shows a schematic of the polypeptide predicted from the sequence (SEQ ID NO:9) of the transcript (Fru#1) schematized in Fig. 7D.

Figure 9 shows the DNA sequence (SEQ ID NO:9) of the transcript (Fru#1) schematized in Fig. 7D.

15 **BRIEF DESCRIPTION OF THE SEQUENCES**

SEQ ID NO:1 is the nucleotide sequence of the 3x *dsx* repeat DNA probe.

SEQ ID NO:2 is the nucleotide sequence of the sense *dsx* repeat 21-mer oligonucleotide.

SEQ ID NO:3 is the nucleotide sequence of the antisense *dsx* repeat 21-mer oligonucleotide.

20 SEQ ID NO:4 is the nucleotide sequence of the -20 sequencing primer.

SEQ ID NO:5 is the nucleotide sequence of the *fru-1* primer.

SEQ ID NO:6 is the nucleotide sequence of the *fru-2* primer.

SEQ ID NO:7 is the nucleotide sequence of the *fru-5-rev* primer.

SEQ ID NO:8 is the nucleotide sequence of the *fru-4-rev* primer.

25 SEQ ID NO:9 is the nucleotide sequence of the Fru#1 cDNA transcript.

SEQ ID NO:10 is the translated amino acid sequence of SEQ ID NO:9.

SEQ ID NO:11 is the nucleotide sequence of the ~ 600 bp *EcoRI* *fru* genomic clone insert containing 3 *dsx* repeats.

30 SEQ ID NO:12 is the nucleotide sequence of the 3' end of the fruitless transcript schematized in Fig. 7E.

SEQ ID NO:13 is the translated amino acid sequence of SEQ ID NO:12.

SEQ ID NO:14 is the expected nucleotide sequence of the fruitless transcript schematized in Fig. 7E.

SEQ ID NO:15 is the translated amino acid sequence of SEQ ID NO:15.

**DETAILED DESCRIPTION OF THE INVENTION****Definitions**

A FRU polynucleotide is defined herein as a polynucleotide that selectively hybridizes with a probe directed to unique sequences in the *fru* polynucleotides presented herein (e.g.,

5 SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:14). Such unique sequences are sequences that do not overlap common regions of other transcription factors, such as the BTB region and zinc (Zn) finger domains. For example, a probe containing the sequence between positions 1870 and 2080 of SEQ ID NO:9 is directed to unique sequences in the *fru* polynucleotides presented herein.

10 A FRU polypeptide is defined herein as a polypeptide encoded by the open reading frame of a FRU polynucleotide.

Regulatory sequences, or control sequences, refer to specific sequences at the 5' and 3' ends of eukaryotic genes which may be involved in the control of transcription. For example, most eukaryotic genes have an AT-rich region located approximately 25 to 30 bases 15 upstream from the site where transcription initiation site. Similarly, most eukaryotic genes have a CXCAAT region (X may be any nucleotide) 70 to 80 bases upstream from the start of transcription.

20 The term "operably linked", as used herein, denotes a relationship between a regulatory region (typically a promoter element, but may include an enhancer element) and the coding region of a gene, whereby the transcription of the coding region is under the control of the regulatory region.

25 A polunucleotide or polypeptide is "derived from" a particular organism if that polunucleotide or polypeptide was originally isolated from that organism. For example, a polynucleotide in a plasmid propagated in *E. coli* is derived from *Drosophila* if that polynucleotide was originally isolated from *Drosophila* mRNA, genomic DNA or cDNA. Alternatively, a polunucleotide or polypeptide is "derived from" a particular organism if the sequence of that polynucleotide or polypeptide is based on the sequence of the corresponding sequence from that organism. For example, a polypeptide is derived from *Drosophila* if the sequence of the polypeptide is the same as the sequence of the corresponding native 30 *Drosophila* polypeptide.

**I. Overview of the Invention**

In the fruit fly *Drosophila melanogaster*, as in other animals, one of the most obvious differences between adults of different sexes are the sex-specific behaviors involved in

reproduction. In flies, reproductive behaviors for males include the detection of females, precopulatory courtship, and finally copulation (for review: Speith, 1974).

Many aspects of reproductive behavior are controlled by the central nervous system (CNS), and may accordingly have a neuronal cell basis. Sexually dimorphic neurons in the 5 CNS are intimately associated with the performance of sex-specific behaviors. In the nervous system, neuronal differences may be manifested in a variety of ways. Neurons may be unique to one sex, or neurons may be present in both sexes but differ in size, shape, anatomical connections, or physiology.

In insects, a variety of sex-specific differences in the CNS have been described both in 10 the sensory integration and in motor output systems. For example, sexually dimorphic sensory input from the moth's male-specific antennal sensory neurons, which detect the air-borne female pheromone, has been shown to form specialized connections only with male-specific interneurons in the antennal lobe (Matsumoto and Hildebrand, 1981). Effector organs, such as genital muscles or internal reproductive organs, are often sex-limited, leading 15 to the establishment of segment specific cohorts of motorneurons, as found for example in the abdominal ganglia of moths (Giebultowicz and Truman, 1984; Thorn and Truman, 1989).

In *Drosophila* certain elements of this species' central and peripheral nervous system, as 20 well as some genital and abdominal muscles, are known to be different in developing or adult males vs. females (Technau, 1984; Lawrence and Johnston, 1986; Stocker and Gendre, 1988; Taylor 1989a,b; Possidente and Murphey, 1989; Taylor and Truman, 1992, Taylor, 1993). However, information regarding the neuronal basis for adult sexually dimorphic behaviors has lagged behind the descriptions of such behaviors and their modification by experience or 25 various mutant genotypes.

Somatic sexual differentiation in the fruit fly *Drosophila melanogaster* is controlled by a 25 genetic regulatory hierarchy that involves the interactions of a number of genes including *Sex-lethal (Sxl)* *transformer (tra)*, *transformer-2 (tra-2)* and *doublesex (dsx)*. Each of these genes has been cloned and characterized at the molecular level. Results of these analyses have revealed that the genes function in a cascade of alternative message RNA (mRNA) processing decisions. An effect of this cascade is the production of sex-specific *dsx* proteins that 30 function as transcriptional regulators that control expression of genes involved in sexual differentiation.

Experiments performed in support of the present invention and described below suggest 35 that *fru* is a member of the *Drosophila* sex-determination regulatory hierarchy and is the first gene unique to a previously unrecognized branch of this hierarchy that governs many aspects of male sexual behavior. These experiments have resulted in the elucidation of the nucleotide

sequence of portions of the *fru* locus in *Drosophila* and cDNA transcripts derived therefrom. According to the teachings presented below, this locus may be an important point in the regulatory hierarchy controlling sexual differentiation in *Drosophila*. Homologous genes in other organisms may play corresponding roles in the sexual differentiation of those  
5 organisms.

As is described more fully below, methods and compositions of the present invention may be used in a variety of ways by one of skill in the art having the benefit of the present disclosure. For example, methods of the present invention may be used to alter the sexual or reproductive behavior of an organism, and/or to identify compounds effective to alter such  
10 behavior. One application of such an alteration in sexual or reproductive behavior is pest control, *e.g.*, insect control.

## II. Role of *fru* in *Drosophila* Sexual Differentiation

In *D. melanogaster*, all aspects of sexual differentiation are controlled by a single  
15 regulatory hierarchy (reviewed by, for example, Wolfner, 1988; Baker, 1989; Cline, 1988; Hodgkin; 1990; Slee and Bownes, 1990; McKeown and Madigan, 1992). The reference of Harry, *et al.*, (1992), discusses these studies against a background of sex-determination genetics in vertebrates. The hierarchy is comprised of an initial series of steps that are concerned with the determination and establishment of sex. After this point, according to the  
20 teachings presented herein, the hierarchy splits into two branches, as is illustrated in Figure 1. The *dsx* branch is established in the literature, while the *fru* branch is based on the results of experiments performed in support of the present invention. The diagram is provided herein as a reference for discussions relating to the possible interactions of other genes and gene products with the methods and compositions of the present invention. The diagram does  
25 not necessarily constitute a mechanistic basis for the functioning of the present invention.

A line in the diagram extending from a gene indicates that it is expressed and has an effect on a downstream gene. If the line ends in an arrow the effect is positive; if it ends in a bar the effect is negative. The activity of genes necessary for female development is on the left and for males is on the right. Results of experiments performed in support of the present  
30 invention suggest that the action of *tra* and *tra-2* may be to cause the *fru* pre-mRNA to be spliced into a non-functional product in females. In the absence of these activities in males, the *fru* pre-mRNA may be spliced into a functional product that is important for the expression of male-specific structures and behaviors.

The initial series of steps in the sex determination hierarchy act to assess the X  
35 chromosome to Autosome ratio (X:A ratio), which is the primary determinant of sex

(Bridges, 1921), and to set the activity of *Sex-lethal (Sxl)*, a master regulatory gene at the top of the hierarchy, to "on" in females and "off" in males (reviewed by, for example, Wolfner, 1988; Baker, 1989; Cline, 1988; Hodgkin, 1990; Slee and Bownes, 1990; McKeown and Madigan, 1992). Once expression of *Sxl* is initiated in females it is maintained "on" by a 5 positive autoregulatory feedback loop in which SXL protein directs the processing of its own pre-mRNA so as to generate a mRNA that encodes SXL protein (see, e.g., the reviews cited above). In males, *Sxl* pre-mRNA is spliced in the default mode which results in the inclusion of a male-specific exon containing stop codons, and hence the male-specific mRNA has no 10 open reading frame.

10 In addition to regulating the processing of its own pre-mRNA the SXL protein also functions in females to control the activity of two subservient branches to the sexual differentiation hierarchy. One of these branches governs somatic sexual differentiation (see above reviews) and the other dosage compensation (review: Lucchesi and Manning, 1987). To regulate somatic sexual differentiation SXL directs the processing of the pre-mRNA of the 15 *transformer (tra)* gene in females so as generate an mRNA with an open reading frame that encodes the TRA protein (Boggs, *et al.*, 1987; Nagoshi *et al.*, 1988). In males, where SXL protein is absent, the *tra* pre-mRNA is spliced by a default pathway, which results in the inclusion of exonic sequences that contain stop codons and hence prevent the synthesis of TRA protein.

20 In females, the TRA protein (which is female-specific), together with the TRA-2 protein (which is made in both sexes), function to regulate the splicing of the pre-mRNA of the *dsx* gene to generate a female-specific *dsx* mRNA (Burtis and Baker, 1989; Nagoshi, *et al.*, 1988; Hedley and Maniatis, 1991; Hoshijima, *et al.*, 1991; Ryner and Baker, 1991). In males, where *tra* protein is absent, the housekeeping splicing machinery carries out the 25 default pattern of *dsx* pre-mRNA processing to generate the male-specific *dsx* pre-mRNA. Both the male- and female-specific *dsx* mRNAs encode Zn-finger transcription factors, which have identical DNA binding domains, but different carboxy termini. The *dsx* gene appears to be the last sex-determination regulatory gene in this branch of the hierarchy, since its proteins have been shown to directly interact with the enhancer sequences of at least one of the genes 30 encoding a terminal sexual differentiation function (Burtis, *et al.*, 1991).

One aspect of sexual differentiation, the formation of the Muscle of Lawrence (MOL), does not appear to be controlled by *dsx*, but is regulated by *tra* and *tra-2* (Taylor, 1992). Results of experiments performed in support of the present invention suggest that the gene immediately below *tra* and *tra-2* in this branch of the hierarchy may be the *fruitless* gene. In 35 particular, the results suggest that the *fru* gene may be negatively controlled by *tra* and *tra-2*.

in females (*i.e.*, the TRA and TRA-2 proteins direct the processing of *fru* pre-mRNA into an mRNA that does not encode a functional product in females); whereas the default pattern of *fru* pre-mRNA processing (which occurs in males) may produce an mRNA encoding functional *fru* product.

5       Based on the phenotypes of extant *fru* alleles, the *fru* branch of the somatic sex determination hierarchy is responsible for the differentiation of the MOL and for expression of normal male courtship behavior. Since both of these phenotypes are determined by the genotype of the nervous system (cf. Siegel *et al.*, 1984, Lawrence and Johnston, 1986), the function of the *fru* branch may be to control at least some aspects of the differentiation of the  
10 CNS, including those responsible for male sexual behavior, and may control other aspects of sexual differentiation. The proposed *fru* branch may also be required to maintain aspects of sexual differentiation in adult organisms, since normal sexual behavior requires continuous wild type *tra-2* function in the adult (Belote and Baker, 1987).

15       Mutations in the *fruitless* locus have striking effects on male courtship behavior: *fru* mutant males initiate courtship of males and females indiscriminately, and are sterile because they are unable to carry out later steps in courtship. Mutations in the *fruitless* gene affect only males, where their most salient phenotype is that they cause males to initiate courtship with both males and females with equal likelihood.

20       III.       FRU Polynucleotides

A.       Molecular Cloning of the *Drosophila fru* Locus

25       DNA sequences corresponding to the *fru* locus in *Drosophila* were isolated in the course of experiments conducted in support of the present invention. A hybridization probe was designed to isolate *fru* sequences based on the discovery, disclosed herein, that the *dsx* and *fru* genes are regulated by a common factor. The probe, which contains three copies of a 13 nucleotide (nt) regulatory sequence repeated six times in the *dsx* transcript, was used to screen a *Drosophila* genomic library as detailed in Example 1. The design and synthesis of the probe are described below in Example 1A - "Generation of Hybridization Probe".

30       Selective hybridization conditions for the probe were determined (Example 1B - "Selective Hybridization Conditions"), and the probe was used to screen a *Drosophila* genomic library (Example 1C - "Genomic DNA Library Screen"). Four clones that were good candidates for DNAs containing multiple copies of the 13 nucleotide *dsx* repeat were isolated (Example 1D - "Southern Blot Analysis of Positive Clones"). The hybridizing fragment from one of these was subcloned into a "BLUESCRIPT SK" phagemid (Stratagene, La Jolla, CA) and the clone (pSK(+)11-R) was sequenced. The sequence is presented herein

as SEQ ID NO:9, and reveals that the insert contained three copies of the 13 nucleotide repeat.

The clone was further characterized as described in Example 2, and was found to: (i) produce sex-specific transcripts, (ii) reside at cytological location 91B, and (iii) fall within a 5 genomic walk that spans over 100 kbp of the *fruitless (fru)* gene.

#### **B. Isolation of *fru* cDNAs**

Example 3, below, details an application of the polymerase chain reaction (PCR; Mullis, Mullis, *et al.*) to obtain the 3' ends of *fru* cDNA transcripts from male and female mRNA 10 (Example 3A - "RACE PCR"). The isolated RACE products were used to design additional PCR primers, which were employed in nested PCR reactions of cDNA to assay for the presence of *fru* transcripts. The primers used to detect these transcript were used in a preliminary screen to identify a *Drosophila* cDNA library containing *fru* transcripts (Example 3B - "Sex-Specific PCR"). A cDNA library thus identified (a λZAP adult heads cDNA 15 library) was then screened for cDNA clones (Example 3C - "cDNA Library Screen"). Nineteen different *fru* cDNAs falling into at least 5 different classes (differing through alternative RNA processing) were isolated from this library, and were characterized to determine how they related to each other and to genomic DNA from the region. The results of this characterization are schematized in Figs. 7D, 7E, 7F, 7G and 7H. The full consensus 20 sequence of one of the transcripts (Fru#1) was determined (SEQ ID NO:9), and is shown in Fig. 9. The consensus sequence of the 3' end of the transcript shown in Fig. 7E (Fru#2) was also determined, and is presented herein as SEQ ID NO:12. Based on extensive Southern mapping, PCR and restriction enzyme analyses, the 5' end of Fru#2 appears identical to that of Fru#1. The sequences diverge at nucleotide number 3012 of Fru#1 (SEQ ID NO:9), 25 corresponding to amino acid residue 503 of the Fru#1 polypeptide (SEQ ID NO:10). The expected full-length nucleotide sequence of Fru#2 is presented herein as SEQ ID NO:14; the corresponding amino acid sequence is presented as SEQ ID NO:15.

#### **C. Isolation of Homologous Sequences from Other Organisms**

FRU polynucleotide sequences of the present invention may be used to isolate 30 homologous sequences from other species, including other insects and mammals. In particular, the FRU polynucleotide sequences may be used to isolate corresponding sequences from insects belonging to the phylum Arthropoda (Arthropods), and more particularly, the 35 order Diptera (flies). Examples of Arthropods from which corresponding sequences may be isolated include fruit flies, such as medflies and mexican, mediterranean, oriental, and olive

fruit flies (for example, other *Drosophila* species (sp.), *Rhagoletis* sp., *Ceratitis* sp. (e.g., *Ceratitis capitata*) and *Dasus* sp. (e.g., *Dasus oleae*)), tse-tse flies, such as *Glossina* sp. (e.g., *Glossina palpalis*), sand flies, such as *Phlebo* sp. (e.g., *Phlebo tomus*)), blowflies, flesh flies, face flies, houseflies, screw worm-flies, stable flies, mosquitos, northern cattle grub and the like.

Several strategies may be pursued to this end. For example, Southern blots containing DNAs from target species may be probed with a portion of the *fru* sequence disclosed herein using a series of hybridization conditions to identify those conditions resulting in selective hybridization. An example of how selective hybridization conditions may be experimentally determined is provided in Example 1B. The screen may be conducted with a series of probes (e.g., ~8 probes, each about 250 bp in length) that span the known *Drosophila fru* sequences.

Effective probes preferably correspond to sequences that are conserved between different species (i.e., coding sequences), and that are not homologous to a large number of non-FRU polypeptides, such as other transcription factors. To this end, portions of the *fru* coding sequence may be used to search DNA databases, and those regions resulting in a minimal number of homologous "hits" to undesired sequences, such as other transcription factors, may be used as cross-species probes. For example, the sequence between positions 1870 and 2080 of the Fru#1 cDNA (SEQ ID NO:9) is not highly homologous to other sequences present in the DNA databases. Probes derived from this region may be effective at isolating *fru* homologs from other species.

Alternatively, Northern blots may be screened with a cDNA probe as described above to identify species which may contain *fru* homolog transcripts. Conditions for selective hybridization may be determined experimentally (e.g., as described in Example 2).

Once selective hybridization conditions are determined, genomic DNA and/or cDNA libraries from the target species are screened to isolate *fru* homolog DNA fragments. The fragments may be sequenced and the sequences arranged into a consensus sequence spanning the *fru* homolog region. Alternatively, the sequences may be used as probes for additional screening, extended using RACE PCR approaches (e.g., as in Example 1), and/or used, in combination with sequences disclosed herein, to design degenerate PCR primers for finding *fru* cognates in yet more distantly related species.

Sequences identified in other species can likewise be used as probes, for example, against genomic and cDNA libraries from that species, to identify the entire genetic locus in that species.

**D. Use of FRU Polynucleotides**

Polynucleotides of the present invention may be used in a screen for compounds effective to alter the sexual or reproductive behavior of an animal, such as a pest insect. Such a screen may include a reporter gene construct in an expression vector. An expression vector 5 bearing a selectable marker can be constructed with a reporter gene (such as chloramphenicol acetyl-transferase acetyl transferase (CAT),  $\beta$ -galactosidase or luciferase) under the control of, for example, a *fru* promoter element, and transfected into a selected host cell (for example, Schneider's Line 2 cells or *Drosophila* Kc cells (Schneider, Ryner and Baker, Hoshijima, K., *et al.*)). After transfection, effects of test compounds on transcription may be 10 measured by the activity of the reporter gene (e.g. CAT) in, for example, crude cell extracts.

Using FRU probes, non-coding regulatory regions adjacent the FRU coding sequences can be derived from genomic DNA samples, for example, from the  $\lambda$ Charon 4A *Drosophila* genomic library. Using FRU specific primers, both the three and five prime ends of the gene are isolated using the PCR rapid amplification of cDNA ends (PCR-RACE) reaction 15 (Frohman, 1988, 1990). Such 5' non-coding regulatory regions contiguous to 5' FRU coding sequences can be fused to reporter genes such that the reporter gene is in-frame with respect to the location of FRU coding sequences. These reporter constructs can then be transformed into a selected host cell.

Reporter gene systems are well known in the art (see, for example, Ausubel, *et al.*). Cell 20 lines and vectors used in reporter gene assays are commercially available (for example, Stratagene, La Jolla, CA; Clontech Laboratories, Palo Alto, CA; Promega Corporation, Madison, WI; American Type Culture Collection, 12301 Parklawn Dr., Rockville MD 20852). One example of a family of commercially-available reporter plasmids are the 25 "pCAT" plasmid (Promega Corp., Madison, WI), that contain a CAT transcription unit and an ampicillin resistance gene.

Candidate compounds can be obtained from a number of sources, including but not limited to, the following. Many pharmaceutical and agrochemical companies have extensive 30 libraries of chemical and/or biological mixtures, often fungal, bacterial, or algal extracts, that would be desirable to screen with the assay of the present invention. Such compounds, or molecules, may be either biological or synthetic organic compounds, or even inorganic compounds.

Transfected cells are treated with a selected compound, and the levels of reporter gene product present in treated and untreated cells is determined and compared. Compounds that result in decreased expression of the reporter gene in treated cells are identified as potentially 35 useful sexual behavior-altering compounds. Alternatively, in the case of reporter systems that

do not kill or substantially alter the cells, the level of reporter expression may be assayed in the same batch of cells both before (basal level) and after treatment. Levels of expression are compared, and a compound is identified as effective if it significantly depresses the level of expression (relative to the basal level) following treatment.

5 It will be appreciated that compounds identified as effective in the cells from one species of a group (e.g., insects) may also be effective in other species of that group. In particular, compounds identified as effective in a model system using cells from one species may be tested as described below for effects on other, related species.

10 Compounds identified by the above screen(s) as potentially effective may be further tested for their ability to alter the sexual or reproductive behavior of a selected organism. For example, a compound identified by the above method may be administered to an insect population to determine if the compound is effective at reducing the reproductive rate of the population.

15 A variety of insects may be targeted by methods of the present invention. For example, insects belonging to the phylum Arthropoda (Arthropods), and more particularly, the order Diptera (flies) are particularly suitable for targeting by the methods of the present invention. Specific examples of Arthropods which may be targeted include fruit flies, such as medflies and mexican, mediterranean, oriental, and olive fruit flies (for example, *Drosophila* species (sp.), *Rhagoletis* sp., *Ceratitis* sp. (e.g., *Ceratitis capitata*) and *Dasus* sp. (e.g., *Dasus oleae*)), tse-tse flies, such as *Glossina* sp. (e.g., *Glossina palpalis*), sand flies, such as *Phlebo* sp. (e.g., *Phlebo tomus*)), blowflies, flesh flies, face flies, houseflies, screw worm-flies, stable flies, mosquitos, northern cattle grubs and the like.

#### IV. FRU Polypeptides

##### 25 A. Production of Recombinant Polypeptides

Polynucleotide sequences of the present invention may be cloned into an expression plasmid, such as p-GEX, to produce corresponding polypeptides. The plasmid pGEX (Smith, *et al.*, 1988) and its derivatives express the polypeptide sequences of a cloned insert fused in-frame with glutathione-S-transferase. Recombinant pGEX plasmids can be transformed into appropriate strains of *E. coli* and fusion protein production can be induced by the addition of IPTG (isopropyl-thio galactopyranoside). Solubilized recombinant fusion protein can then be purified from cell lysates of the induced cultures using glutathione agarose affinity chromatography according to standard methods (Ausubel, *et al.*).

35 Affinity chromatography may also be employed for isolating  $\beta$ -galactosidase fusion proteins (such as those produced by lambda gt11 clones). The fused protein is isolated by

passing cell lysis material over a solid support having surface-bound anti- $\beta$ -galactosidase antibody.

Isolated recombinant polypeptides produced as described above may be purified by standard protein purification procedures. These procedures may include differential 5 precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis and affinity chromatography.

In addition to recombinant methods, FRU proteins or polypeptides can be isolated from selected cells by affinity-based methods, such as by using anti-FRU antibodies (described below). Further, FRU peptides may be chemically synthesized using methods known to these 10 skilled in the art.

#### **B. Use of FRU Polypeptides**

Polypeptides of the present invention may be used in a number of ways, including the generation of antibodies. The polypeptides may be used in unmodified form, or they may be 15 coupled to appropriate carrier molecules, such as bovine serum albumin (BSA) or Keyhole Lympet Hemocyanin (KLH) (available from, for example, Pierce, Rockford, IL).

To prepare antibodies, a host animal, such as a rabbit, is typically immunized with the purified polypeptide or fusion protein (generated using, for example glutathione-S-transferase as described above). The host serum or plasma is collected following an appropriate time 20 interval, and the serum is tested for antibodies specific against the polypeptide.

The gamma globulin fraction or the IgG antibodies of immunized animals can be obtained, for example, by use of saturated ammonium sulfate precipitation or DEAE Sephadex chromatography, affinity chromatography, or other techniques known to those skilled in the art for producing polyclonal antibodies.

25 Alternatively, purified antigenic polypeptide or fused antigen protein may be used for producing monoclonal antibodies. In this case, the spleen or lymphocytes from an immunized animal are removed and immortalized or used to prepare hybridomas by methods known to those skilled in the art (see, *e.g.*, Harlow, *et al.*). Antibodies secreted by the immortalized cells are screened (see, *e.g.*, using enzyme linked immunosorbent assay 30 (ELISA) or a Western blot) to determine the clones that secrete antibodies of the desired specificity (see, *e.g.*, Ausubel, *et al.*).

Antibodies generated as described above may be used in a variety of ways. For example, 35 antibodies generated against FRU polypeptides may be used in salivary glands to identify the chromosomal locations to which the FRU protein binds on the giant polytene chromosomes of these cells. The resolution available with this technique is such that it is typically possible to

ascertain within a few tens of kb where the protein is binding. This enables a relatively rapid identification of the gene in question by determining which genes in the region are expressed in a spatial and temporal pattern consistent with present knowledge of *fru* expression and male courtship behavior. This approach may also be used in screens of other insects with 5 polytene chromosomes to identify FRU polypeptide targets in those species.

Alternatively, DNA sequences to which the FRU polypeptide binds may be identified, for example, by employing anti-FRU antibodies in DNA/protein interaction assays. Restriction enzyme-digested DNA may be combined with purified FRU protein (and optionally, nuclear extracts from the cells of interest) and size fractionated in duplicate (one preparatory, one 10 analytical) lanes on a polyacrylamide gel. Material from the analytical lane may be blotted and probed with an anti-FRU antibody to determine the location of a FRU-DNA complex in the gel. The complex may then be excised from the corresponding preparatory lane of the gel, and the DNA contained therein may be isolated and cloned for further analysis.

DNA sequences to which the FRU polypeptide binds may be used to identify targets for 15 pest control screens. For example, the approach may be used to identify gene products involved in sexual recognition (distinguishing males from females). This process is thought to involve the reception of pheromone cues by receptors. Genes for such receptors may be targets of regulation by FRU gene products. Identification of pheromone receptors in insects may be used to screen for compounds which affect the functioning of those receptors. Such 20 compounds may find wide application in the area of insect control.

Alternatively, recombinant FRU polypeptides may be labeled (e.g., with  $^{125}\text{I}$ ) and used in a screen such as is outlined above to identify DNA fragment that bind the polypeptides. The location of the labeled protein in the blot is determined directly, without the use of an anti-FRU antibody, and corresponding DNA sequences are similarly isolated. DNA sequences 25 identified by any of the methods described above may be used to screen for compounds that interfere with the binding of FRU protein to its target DNA, using screens similar to that described above for the screening of compounds that interfere with the transcriptional activation of *fru*.

Antibodies generated as described above may also be used to co-immunoprecipitate 30 proteins which interact with FRU polypeptides (partners of FRU). Partners of FRU may be involved in sex-specific or non-sex-specific functions, but the identification of such partners may result in the isolation of new genes involved in sex behavior and/or viability of flies and other insects.

Partners of FRU may also be isolated using, for example, the yeast two-hybrid system. 35 The presence of a BTB domain in FRU polypeptides suggests that the polypeptides are

involved in protein-protein interactions. The two hybrid system may be used to isolate polypeptides that interact with FRU polypeptides.

Two hybrid protein interaction assay methods (two hybrid protein-protein interaction screens) provide a simple and sensitive means to detect the interaction between two proteins in living cells. The assays are based on the finding that most eukaryotic transcription activators are modular (*e.g.*, Brent, *et al.*), *i.e.*, that the activators typically contain activation domains that activate transcription, and DNA binding domains that localize the activator to the appropriate region of a DNA molecule.

In a two hybrid system, a first fusion protein contains one of a pair of interacting proteins fused to a DNA binding domain, and a second fusion protein contains the other of a pair of interacting proteins fused to a transcription activation domain. The two fusion proteins are independently expressed in the same cell, and interaction between the "interacting protein" portions of the fusions reconstitute the function of the transcription activation factor, which is detected by activation of transcription of a reporter gene.

At least two different cell-based two hybrid protein-protein interaction assay systems have been used to assess binding interactions and/or to identify interacting proteins. Both employ a pair of fusion hybrid proteins, where one of the pair contains a first of two "interacting" proteins fused to a transcription activation domain of a transcription activating factor, and the other of the pair contains a second of two "interacting" proteins fused to a DNA binding domain of a transcription activating factor.

The yeast GAL4 two hybrid system (Fields, *et al.*; Chien, *et al.*; Durfee, *et al.*; Bartel, *et al.*) was developed to detect protein-protein interaction based on the reconstitution of function of GAL4, a transcriptional activator from yeast, by activation of a *GAL1-lacZ* reporter gene. Like several other transcription activating factors, the GAL4 protein contains two distinct domains, a DNA binding domain and a transcription activation domain. Each domain can be independently expressed as a portion of a fusion protein composed of the domain, and a second, "bait" interacting protein. The two fusion proteins are then independently expressed together in a cell. When the two GAL4 domains are brought together by a binding interaction between the two "interacting" proteins, transcription of a reporter gene under the transcriptional control of GAL4 is initiated. The reporter gene typically has a promoter containing GAL4 protein binding sites (GAL upstream activating sequences, UAS<sub>G</sub>).

In one example of the use of a two hybrid system to isolate partner(s) of FRU, a FRU polypeptide is fused to the GAL4 DNA binding domain (G4BD) in a yeast expression vector (pG4AD-FRU). The vector is used to generate yeast cells harboring pG4AD-FRU and a

GAL4-activated reporter gene (e.g., LacZ), which are then transformed with one of three fusion libraries. Each library carries fusions between the transcription activating domain of yeast GAL4 (G4AD) and insect (e.g., *Drosophila*) genomic DNA restriction enzyme fragments (e.g., *Sau*3A1 fragments) in one of the three reading frames.

5 The yeast cells containing the libraries are screened (e.g., using a  $\beta$ -galactosidase ( $\beta$ -gal) assay on plates containing the chromogenic substrate X-gal) for expression of the reporter. Reporter-expressing cells are identified as possibly containing *Sau*3A1 DNA fragments encoding polypeptides capable of interacting with the FRU polypeptide.

10 A second two hybrid system, described in detail in Ausubel, *et al.*, utilizes a native *E. coli* LexA repressor protein, which binds tightly to appropriate operators. A plasmid is used to express one of a pair of interacting proteins (the "bait" protein) as a fusion to LexA.

15 The plasmid expressing the LexA-fused bait protein is used to transform a reporter strain of yeast, such as EGY48. In this strain, binding sites for LexA are located upstream of two reporter genes. In the first reporter system, the upstream activation sequences of the chromosomal LEU2 gene--required in the biosynthetic pathway for leucine (Leu)--are replaced in EGY48 with lexA operators, permitting selection for viability when cells are plated on medium lacking Leu. In the second reporter system, EGY48 harbors a plasmid, pSH18-34, that contains a lexA operator-lacZ fusion gene, permitting discrimination based on color when the yeast is grown on medium containing Xgal (Ausubel, *et al.*).

20 LexA and GAL4 each have different properties that should be considered when selecting a system. LexA is derived from a heterologous organism, has no known effect on the growth of yeast, possesses no residual transcriptional activity, can be used in GAL4<sup>+</sup> yeast, and can be used with a Gal-inducible promoter. Because GAL4 is an important yeast transcriptional activator, experiments must be performed in *gal4* yeast strains to avoid background from 25 endogenous GAL4 activating the reporter system. Both two hybrid systems have been successfully used for isolating genes encoding proteins that bind a target protein and as simple protein binding assays (see, e.g., Yang, *et al.*, Gyuris, *et al.*), and both can be applied to the identification of polypeptides that interact with the FRU polypeptide.

30 V. Generation of New *Fru* Phenotypes

Modified *fru* constructs may be reintroduced into flies to generate *Fru* alleles with dominant behavioral and/or sterility phenotypes. Such constructs include those in which either the DNA binding domain or the N-terminal BTB domain are truncated, as well as constructs that ectopically express *fru* cDNAs under a ubiquitous (e.g., hsp70) promoter.

While the presently-known alleles of *fru* are recessive, many loci in *Drosophila* have both dominant and recessive alleles. One such locus, *doublesex* (Baker and Ridge, 1980), is also involved in the regulatory hierarchy controlling sexual differentiation and is a Zn finger-containing transcription factor (Burtis and Baker, 1989).

5 Constructs effective at conferring dominant sterile phenotypes may be engineered into vectors suitable for transforming other types of insects, such as insects belonging to the phylum Arthropoda (Arthropods), and more particularly, the order Diptera (flies). Specific examples of Arthropods which may be transformed include flies, such as medflies and mexican, mediterranean, oriental, and olive fruit flies (for example, *Drosophila* species (sp.), 10 *Rhagoletis* sp., *Ceratitis* sp. (e.g., *Ceratitis capitata*) and *Dasus* sp. (e.g., *Dasus oleae*)), tse-tse flies, such as *Glossina* sp. (e.g., *Glossina palpalis*), sand flies, such as *Phlebo* sp. (e.g., *Phlebo tomus*), blowflies, flesh flies, face flies, houseflies, screw worm-flies, stable flies, mosquitos, northern cattle grub and other pests.

Such transgenic insects have been made by injecting a vector containing cloned DNA and 15 a selectable marker into embryos and selecting transgenic progeny (Miller, *et al.*). Mutant insects produced in this manner may be grown and used in sterile-release programs to aid in controlling pest insect populations. Such programs have been demonstrated to be successful in controlling insect pest populations (see, for example, Wong, *et al.*, Calkins, *et al.*).

20 Specimens made sterile by the introduction of a dominant mutation of *Fru* or its homologs offer an advantage in that the sterility gene is propagated through a series of generations by females carrying the mutation mating with wild-type males. Of course, the sterile males also aid in reducing the population by (fruitlessly) courting both wild-type males and females.

25 The following examples illustrate but in no way are intended to limit the present invention.

#### MATERIALS AND METHODS

Unless indicated otherwise, chemicals and reagents were obtained from Sigma Chemical 30 Company (St. Louis, MO) or Mallinckrodt Specialty Chemicals (Chesterfield, MO), restriction endonucleases were obtained from New England BioLabs (Beverly, MA), and other modifying enzymes and biochemicals were obtained from Pharmacia Biotech (Piscataway, NJ), Boehringer Mannheim (Indianapolis, IN) or Promega Corporation (Madison, WI). Materials for media for cell culture were obtained from Gibco/BRL 35 (Gaithersburg, MD) or DIFCO (Detroit, MI). Unless otherwise indicated, manipulations of

*Drosophila*, cells, bacteria and nucleic acids were performed using standard methods and protocols (see, e.g., Ashburner; Sambrook, *et al.*; Ausubel, *et al.*).

#### EXAMPLE 1

5

##### Molecular Cloning of the *fru* Gene Locus

###### A. Generation of Hybridization Probe

A DNA probe (SEQ ID NO:1) containing 3 copies of the *dsx* 13 nucleotide (nt) repeated sequence was generated as follows. Two 21 nucleotide complementary single-stranded (ss) oligonucleotides (SEQ ID NO:2, SEQ ID NO:3) were synthesized by the Pan Facility (Beckman Center B065, Stanford University Medical Center, Stanford, CA).

The oligonucleotides were hybridized to each other by heating a solution containing equimolar amounts of the two oligonucleotides (130  $\mu$ g of each) to 95°C in a heater block, and then removing the block from the heater and allowing it to cool to room temperature over approximately 30 minutes.

15 The resulting double-stranded (ds) DNA fragment contained complementary four base 5' protruding ends. The 5' ends were phosphorylated with 2 mM ATP and 20 units of polynucleotide kinase (New England BioLabs, Beverly, MA) for 2 hours at 37°C. The DNA was then ethanol precipitated and resuspended in 40  $\mu$ l of water.

20 The phosphorylated dsDNA fragment was multimerized using T4 DNA ligase (New England BioLabs) by incubating the whole DNA sample (260  $\mu$ g) in ligation buffer (New England BioLabs) containing 30 units of T4 DNA ligase for 1 hour at 20°C. The reaction mixture was then digested with 100 units of restriction endonucleases *Bam*HI and *Bgl*II (New England BioLabs) for 1 hour under conditions recommended by the manufacturer. This procedure digested molecules ligated together in opposite orientations. Multimers comprised 25 of repeat fragments having the same orientation remained intact. The reaction mixture was then cooled on ice, mixed with gel loading buffer, and the DNA fragment multimers contained therein were size fractionated by agarose gel electrophoresis on a 1.5% gel.

Multimers ranging from about 63 bases to about 126 bases in length were excised from the gel, partially purified by electroelution (Sambrook, *et al.*), and subcloned into the unique 30 *Bam*HI restriction endonuclease site of the phagemid "BLUESCRIPT SK(+)" (Stratagene, La Jolla, CA). The inserts of several clones were sequenced, and an isolate (pSK(+)3XR) containing 3 copies (3 $\times$  repeats) of the synthetic dsDNA fragment was identified. This plasmid was further modified by deleting the region between the *Kpn*I and *Pst*I restriction sites to facilitate a higher level of incorporation of radioactive nucleotides into hybridization 35 probes made from the plasmid.

A single stranded (ss) radioactive probe was generated as follows: ssDNA was obtained from the f1 *ori*-containing pBSK(+)3×R upon co-infection of the host cells with helper phage following manufacturer's instructions (Stratagene). One  $\mu$ g of the ssDNA was combined with 2.5 ng of -20 primer (SEQ ID NO:4), 5 units of Klenow fragment (GIBCO 5 BRL Research Products/Life Technologies, Gaithersburg, MD), 70  $\mu$ Ci each of  $\alpha$ -<sup>32</sup>P-dCTP and  $\alpha$ -<sup>32</sup>P-dATP, and 30  $\mu$ M each dGTP and dTTP cold nucleotides in 30  $\mu$ l of 20 mM Tris-HCl, pH 8.5, 10 mM MgCl<sub>2</sub> buffer to make a labeled complementary copy of the single stranded template (Burtis and Baker, 1989).

The radioactively-labeled insert portion of the plasmid was excised by digestion with *Xba*I and *Bam*HI and was gel purified using low melting-point agarose ("NUSIEVE GTG"; FMC BioProducts, Rockland, Maine). The gel slice containing the probe was melted and added directly to hybridization reactions described below.

**B. Selective Hybridization Conditions**

15 Selective hybridization conditions for library screening were determined as follows. 4  $\mu$ g of total genomic *Drosophila* DNA was digested with *Eco*RI or *Bam*HI, size fractionated by 0.9% agarose gel electrophoresis and transferred to a nylon membrane (Schleicher & Schuell, Keene, NH).

20 The membrane was hybridized overnight with the 3× repeats probe under standard conditions (Sambrook, *et al.*), using 6× SSC, 5× Denhardt's reagent, 0.5% Sodium dodecyl sulfate (SDS), and 100  $\mu$ g/ml denatured and sheered salmon sperm DNA (no formamide) at 42°C. Following hybridization, the filter was washed under the same salt conditions but at increasing temperatures. The results are shown in Figures 2A (47°C final wash) and 2B (51°C final wash). The 47°C wash resulted in detection of several bands in both the *Bam*HI 25 and *Eco*RI digests. Only two prominent fragments were observed in both digests following the 51°C wash. In both digests, one of the fragments is of the size expected for the *dsx*-containing fragment (indicated with arrows), and the other, having a smaller size (~ 600 bp in the *Eco*RI digest and ~ 5 kb in the *Bam*HI digest), is indicated by a "?".

30 These results suggest that the hybridization probe is detecting sequences from two genes - - the *dsx* gene from which it was designed, and a second, unidentified gene.

**C. Genomic DNA Library Screen**

The labeled 3× repeats probe described above was used to screen a lambda Charon 4A (Maniatis, *et al.*, 1978) *Drosophila* genomic library for homologous sequences. As

equivalent of eight genomes' worth of DNA were screened using the conditions described above with a 40°C final wash.

Forty two positive plaques were detected. Eight of these were determined to be from *dsx*. The remaining 34 were isolated and compared with each other using cross-

5 hybridization analysis, which indicated that the 34 non-*dsx* clones represented 12 different sets of clones.

#### D. Southern Blot Analysis of Positive Clones

The clones were further characterized by Southern analysis. One clone from each set 10 was digested with *EcoRI*, size-fractionated on a gel, and blotted onto a nitrocellulose filter. The filter was hybridized with the 3× repeat probe and washed at 40°C as above.

Hybridizing bands were detected by autoradiography (Fig. 3A). The same filter was then hybridized again with a second probe containing 5 copies of the 13 nt repeat sequence (but no other sequence in common with the first probe). The second probe was generated from a 260 15 base-pair (bp) fragment of *dsx* (positions 2793 to 3053; Burtis and Baker, 1989). The filter was washed and subjected to autoradiography as above, and is imaged in Figure 3B.

Four of the clones, indicated in Fig. 3B by "\*", hybridized with both probes and were thus considered to be the best candidates for non-*dsx* DNA containing multiple copies of the 13 nt repeat sequence. One of these (Figs. 3A and 3B, lanes labelled 11), representing eight 20 of the 34 originally-identified non-*dsx* clones, had a particularly strong hybridization signal. This lambda phage clone, termed λCh4A-11, was characterized further as described below.

#### E. Sequence Analysis of a Candidate Clone

Clone λCh4A-11 contained a ~ 600 bp *EcoRI* insert which hybridized to the 3× repeat 25 probe. This fragment was isolated and subcloned into the *EcoRI* site of pBluescript SK(+), generating pSK(+)11-R. Approximately 550 bp of the ~ 600 bp insert of pSK(+)11-R were sequenced using standard dideoxy termination sequencing reactions (Sanger, *et al.*) with a "SEQUENASE 2.0" sequencing kit (United States Biochemical, Cleveland, OH). The sequence (presented in Fig. 4 and as SEQ ID NO:11) revealed that the clone contained 3 30 copies of the 13 nt *dsx* repeat sequence (indicated by boxes in Fig. 4). Also indicated in Figure 4 is the location of the two *EcoRI* sites. Bases whose sequence was not precisely determined are indicated by "N". The seven remaining clones in the set represented by λCh4A-11 also contained the ~ 600 bp *EcoRI* fragment (SEQ ID NO:11) that hybridized strongly to the 3× repeats probe.

EXAMPLE 2Characterization of pSK(+)11-RA. Northern Blot Analysis

To test whether the genomic fragment insert was from a transcription unit, an anti-sense 5 radioactive riboprobe was synthesized from the ~ 600 bp insert of pSK(+)11-R using standard techniques (Sambrook, *et al.*) and used to probe a blot containing poly(A+) male and female RNA from whole adult flies (Figure 5). The sense/antisense orientation of the insert was deduced from a comparison of the 13nt repeat sequence in the clones with the same repeat sequences in *dsx*. The blot was hybridized at 65°C using standard RNA blot 10 hybridization techniques (Sambrook, *et al.*), washed at 40°C, imaged (Fig. 5A), washed at 65°C, and imaged again (Fig. 5B). Imaging was done using autoradiography.

The RNA was isolated using standard methods. Briefly, adult flies were homogenized in 4M guanidium isothiocyanate, 10 mM EDTA, 100 mM Tris pH 7.5 and 1%  $\beta$ -mercaptoethanol, then layered onto a 5.7 M CsCl, 0.1 M EDTA cushion and centrifuged at 15 150,000  $\times$  g for 12 hours. The RNA pellet was then resuspended in 10 mM Tris-HCl pH 15 7.5, 5 mM EDTA and 0.1% sodium dodecyl sulfate (SDS). After phenol extraction and ethanol precipitation the RNA was selected on oligo d(T) cellulose type 7 (Pharmacia, Piscataway, NJ) as described in Sambrook, *et al.*

The images, shown in Figures 5A and 5B, detected the presence of at least 4 transcripts. 20 2 of which (arrows in Figs. 5A and 5B) appeared to be expressed in a sex-specific manner (one in each sex). A ~ 5 kilobase (kbp transcript was expressed in males ("m") and a ~ 6 kbp transcript was detected in females ("f").

B. Chromosomal Localization

25 *In situ* hybridization on squashes of salivary gland polytene chromosomes (Ashburner) was carried out to determine where on the *Drosophila* chromosomes the set of clones represented by clone pSK(+)11-R resides. DNA from 2 of the 8 overlapping lambda phage clones (clones  $\lambda$ Ch4A-11 and  $\lambda$ Ch4A-19) was used to generate biotinylated probes 30 (Ashburner), which were used to probe polytene chromosome squashes using standard methods (Ashburner). The probes hybridized to cytological location 91B, suggesting that the sequences isolated herein may correspond to the *fru* gene, whose locus also resides at 91B. Further evidence linking the clones to the *fru* locus was obtained from results showing specific hybridization of the clones to DNAs obtained during a genomic walk spanning the *fru*-containing region of chromosome 3.

EXAMPLE 3Isolation of *fru* cDNAs

Three different cDNA libraries from *Drosophila melanogaster*, including λnvx male larval and female larval cDNA libraries (obtained from Dr. S. Elledge, Baylor College of Medicine, Houston, TX) and a λgt10 larval disc cDNA library (obtained from Drs. A. Cowman and G. Rubin, University of California, Berkeley, CA), were screened by conventional methods using a probe generated from the insert of clone pSK(+)11-R. However, no *fru* cDNAs were detected in these screens, presumably due to low levels of *fru* expression.

10

A. RACE PCR

Due to the apparent rarity of *fru* mRNA, a 3' end anchored (Frohman, *et al.*) polymerase chain reaction (PCR; Mullis, Mullis, *et al.*) approach was employed to isolate *fru* transcript(s). Two nested primers (fru-1 - SEQ ID NO:5; fru-2 - SEQ ID NO:6) were synthesized as above. The sequences of the primers corresponded to sequences near the 5' end of the pSK9(+)11-R insert. The locations corresponding to the primer sequences are indicated by arrows, labeled as "1" (fru-1) and "2" (fru-2), in Fig. 6A, which shows a schematic of the ~600 bp insert of pSK9(+)11-R. The positions of the 13 nt repeat sequences are shown as black boxes in Fig. 6A.

20 A 3' RACE kit (GIBCO BRL Research Products/LIFE TECHNOLOGIES, Inc., Gaithersburg, MD) was used to generate PCR products from poly (A+) RNA, isolated as described above, from either adult males or adult females. Specific amplification products (~400 bp from male RNA and ~450 bp from female RNA) were detected and determined to contain sequences having homology to the pSK(+)11-R insert by Southern analysis. The 25 PCR products were subcloned and partially sequenced. The sequences corresponded to the sequence near the 5' end of the pSK(+)11-R insert, which appeared to be spliced at a site just downstream of the repeats to different downstream exons. The male- and female-specific 3'RACE products are shown schematically in Figs. 6B and 6C, respectively, in relation to the pSK(+)11-R insert shown in Fig. 6A.

30

B. Sex-Specific PCR

To confirm that the isolated 3' RACE products reflected the structure of authentic *fru* transcripts, new primer sets were synthesized from sequence of the putative male and female PCR products. The positions of these primers are indicated in Figs. 6B and 6C by arrows.

35 The male primer, fru-5-rev, had the sequence represented by SEQ ID NO:7 and the female

primer, fru-4-rev, had the sequence represented as SEQ ID NO:8. These sex-specific primers were paired with fru-1 and fru-2 primers to generate nested primer sets for two rounds of the PCR. The first round was performed with fru-1 and either fru-4-rev or fru-5-rev, and the second round with fru-2 and again with either fru-4-rev or fru-5-rev.

5 These primer sets were used to amplify cDNA generated from several different batches of male- and female-specific poly (A+) RNA. The "female" 3' RACE product, amplified by primers fru-2 (SEQ ID NO:6) and fru-4-rev (SEQ ID NO:8) was subsequently consistently detected in different batches of RNA from both sexes, suggesting that it corresponded to a portion of an authentic *fru* mRNA. Due to the relatively small size of this fragment (450 bp) 10 as compared to the *fru* transcripts detected in Northerns (~ 5-6 kbp; see above), this fragment most likely did not contain a full-length *fru* transcript. To isolate full-length cDNA transcripts, the same primer set (primer fru-2 (SEQ ID NO:6) and fru-4-rev (SEQ ID NO:8) was used in a preliminary screen of a series of *Drosophila* cDNA libraries to identify those 15 libraries which contained *fru* transcripts.

15 Libraries screened included the three listed above plus a λgt10 adult heads cDNA library (obtained from Dr. A. Cowman) and a λZAP (Stratagene, LaJolla, CA) adult heads cDNA library (obtained from Dr. T. Schwarz, Stanford University, Stanford, CA; DiAntonio, *et al.*). The only consistent positive results obtained with the preliminary screen were with the lambda ZAP head cDNA library. Accordingly, this library was screened to isolate *fru* cDNA 20 clones, as described below.

### C. cDNA Library Screen

25 Two-thirds of the complexity of the lambda ZAP head cDNA library described above were screened using conventional methods with labeled "female" 3'RACE product as a probe.

Nine different overlapping cDNAs were isolated. They were characterized by restriction mapping and Southern analysis, including hybridization to the DNAs from the genomic walk, and by cross hybridization to each other. These cDNAs represented at least 3 different 30 classes of transcripts. However, none had the exact structure of the 3' RACE product that was used as the probe to detect them, suggesting that these cDNAs represented only a subset of *fru* transcripts.

Accordingly, the library was rescreened with various portions of the 9 cDNAs. This screen resulted in the identification of 10 new cDNAs that overlapped each other as well as the 9 previously identified cDNAs. Molecular analysis of the new cDNAs revealed two

additional classes of transcripts, including one that contained the sequence found in the "female" 3' RACE product.

A member of each of the five classes was mapped to the DNAs from the genomic walk described above. Fragments from the 5' parts of the cDNA clones mapped to two regions in the distal half of the walk. The 3' end portions of the cDNAs did not hybridize to the walk. The walk was therefore extended in the proximal direction using the cosmid HX1 (obtained from Dr. K. Moses, University of Southern California, Pasadena, CA; Moses, *et al.*), which overlaps the proximal end of the walk. This cosmid was restriction mapped, digested, and blotted for Southern analysis with probes from the 3' end portions of the cDNAs.

Results from the above analyses are shown schematically in Figures 7A, 7B, 7C, 7D, 7E, 7F, 7G and 7H. 5' to 3' is from right to left. Figure 7A shows a schematic of the DNA fragments isolated (f10A, f9A, f3A, f2A, f1D, f1H, f4B, f5C and f7A) as part of a genomic walk spanning the *fru* locus, as well as a schematic of the location of the HX1 cosmid, relative to the map of the *fru* region shown in Fig. 7B. Figure 7B shows a schematic of the *fru* region of chromosome 3, indicating the positions of known *fru* lesions (mutants *fru-2*, *fru-4*, *fru-3* and *fru-1*). The numbers on the scale correspond to kilobases. *fru-1* is depicted by a zig-zag line to indicate an inversion breakpoint, while *fru-2*, *fru-3* and *fru-4* are shown as boxes to indicate insertion of P-element sequences. Figure 7C shows a schematic of two *fru* deficiencies, Df(3R)P14 and Df(3R)ChaM5, relative to the map of the *fru* region shown in Fig. 7B.

Figures 7D, 7E, 7F, 7G and 7H show schematic diagrams of the location of sequences comprising five *fru* cDNA transcripts relative to the map of the *fru* region shown in Fig. 7B. Exons are indicated as boxes and introns as lines. The dark boxes near the 3' ends of the transcripts correspond to exons that contain potential Zn finger sequences, discussed below.

The locations of the 13 nt *dsx* repeats are indicated by "\*".

The results indicate that the 3' ends of the cDNAs correspond to the genomic region spanned by HX1, and demonstrated that *fru* transcripts can contain alternative 3' end exons.

#### D. Sequence Analyses of cDNA Clone Fru#1

One of the isolated cDNAs (shown schematically in Fig. 7D) was sequenced in its entirety. The consensus sequence of this transcript (Fig. 9; SEQ ID NO:9), termed Fru#1, contains one long open reading frame that encodes a 675 amino acid polypeptide (SEQ ID NO:10). The sequence was used to search the Swiss-prot 30 and PIR 42 data bases for homologous sequences (using software from IntelliGenetics Inc., Mt. View, CA). Further, SEQ ID NO:10 was scanned for protein motifs using IntelliGenetics "QUEST" software and

the "PROSITE 12" data bank. These analyses revealed the presence of a highly conserved N-terminal domain, termed BTB domain, found in a number of known transcriptional factors (Zollman, *et al.*), and a single zinc (Zn) finger at the C-terminal of the Fru#1 cDNA (suggesting the presence of a DNA binding domain).

5 A schematic of the Fru#1 polypeptide is shown in Fig. 8. Three copies of the 13 nt repeat sequence are found in the 5' untranslated region just upstream of the ATG initiation codon. The polypeptide contains a BTB domain adjacent the repeats and a Zn finger domain near the C-terminus. The nucleotide sequence of Fru#1 is shown in Fig. 9. The 13 nt repeat regions are underlined, the coding sequence is capitalized, and the ATG initiation codon and  
10 TAA termination codon are in bold.

E. Sequence Analyses of cDNA Clone Fru#2

The 3' portion of the cDNAs shown schematically in Fig. 7E was sequenced as described above. The consensus sequence of the 3' end of this transcript (Fru#2) is presented as SEQ 15 ID NO:12. The 5' end of Fru#2 was analyzed extensively using Southern mapping, PCR and restriction enzyme analyses. The results of these analyses strongly suggest that the sequence of the 5' end of Fru#2 is identical to that of Fru#1. The sequences diverge at nucleotide number 3012 of Fru#1 (SEQ ID NO:9), corresponding to amino acid residue 503 of the Fru#1 polypeptide (SEQ ID NO:10). The expected full-length nucleotide sequence of Fru#2 20 is presented herein as SEQ ID NO:14; the corresponding amino acid sequence is presented as SEQ ID NO:15.

25 While the invention has been described with reference to specific methods and embodiments, it is appreciated that various modifications and changes may be made without departing from the invention.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: The Board of Trustees of the Leland Stanford Junior University  
Board of Reagents, The University of Texas System
- (ii) TITLE OF INVENTION: Methods and Compositions for Altering Sexual Behavior
- (iii) NUMBER OF SEQUENCES: 15
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Dehlinger & Associates
  - (B) STREET: 350 Cambridge Avenue, Suite 250
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94306
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: PCT
  - (B) FILING DATE: 09-FEB-1996
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/386,495
  - (B) FILING DATE: 10-FEB-1995
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Sholtz, Charles K.
  - (B) REGISTRATION NUMBER: 38,615
  - (C) REFERENCE/DOCKET NUMBER: 8600-0153.41
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (415) 324-0880
  - (B) TELEFAX: (415) 324-0960

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 63 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: 3x repeat probe
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GATCCATCTT CAATCAACAT AGATCCATCT TCAATCAACA TAGATCCATC TTCAATCAAC

60

ATA

63

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: sense dsx repeat 21-mer
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GATCCATCTT CAATCAACAT A

21

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: antisense dsx repeat 21-mer
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GATCTATGTT GATTGAAGAT G

21

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: -20 sequencing primer
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTAAAACGAC GGCCAGT

17

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: fru-1 primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GACGTGTGAC GATGGAGCAA C

21

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: fru-2 primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGATCCAGAT CGAAAGAGAA TATCATC

27

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: fru-5 rev primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCTGTCGACA TGCCATAGGT GAATAGGC

28

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: fru-4 rev primer
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGGC GTGATC ATTATGATAT TGTAGCAA

28

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4835 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: Fru#1 cDNA
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1507..3534
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAATTCGGCA CGAGATTCAC	CTATGGCATA TCATCAGCAA	CACACATCAA CGCACTTCTC	60
TGCTATGTCT GCAATCAACC	AAAATATCAA AAAAAAAAAG	AAAAACAAAA AGAGTCAACA	120
TCAATTAA AGTTTTACG	TTGGTTGAA AGAGTTAAA	ATGCCCTTAA CTATTAACGC	180
CCAAAAGTAA ACGTAGATTA	AAGTAATATT AGCCAATCAA	TCGTAAAATA TCAGCTTCG	240
TTTTTTAAAA CTTACCAATG	GACTTGATC CCATCAATTG	CAAATCTAAA GTAGAGAAAT	300
AGAGAGAGAT AAGAGATATA	ATATCACTAA CCAAAAGTGT	TTGCCACGAG TATTAATG	360
TTAACTACTA CAATAGAATA	CGTATTCTTG TTTCTTCGC	TAGTATGTAT AAGCAAAC	420
ACTGCAAGAA ACAACACCAA	CTAATTAATA TTTAATAGCA	TAATGGTAAT ATCGTAAGAA	480
TATCATAGAT TTAAGGCAGA	GCATTTCAGA CAGCACTTGT	ACCGTTCTAG ACTTAAGTAT	540
TCGAAGTATA CGTAACTCAA	GCAATCCAAT AACAAATACT	AAGTAGAAGT TCTTTCAA	600
ATAATACTAT ACACGAATCC	TTCAGTCAAA CCCCCTACAA	TATTACTTAG ATAAACATAT	660
AGTATTATAT AGCCAAAGCC	AGGAAAGGAG TTGTAAGCCA	TTGCATATAT ATATTTGGTA	720
GATAAAGAAC AGCTAACGAA	AGGGTCCACA AGCTACCCAT	AACTTACTTA GAATAACTAA	780
ACACAACTAG CCAAGAAGTA	GATATCTATA TATATATCGA	GTGTTGCTAA CATCAAAGTA	840
TACGTAAATT GAAAACCAAG	AATTTTGCTT AGCTTAAATA	ACACTCTTTC AAAGCAATAC	900
CATAAACAAAT AATTACAAGT	TAACGCAACT AAACACATAT	TGTATACCAAG ATAGTTTATG	960

CCTAAACACT ACTAGTAGCC CTAAGTCCTA GGCATAAACCC GAGCACCAAG GCGAGATATG	1020
CACCCATGTA AAATGCAGAA ATTAATTACC AAGAGTACAA ACTGTAAAGG AAACCCCTAT	1080
TGAAGCTCAA TTGGCCAGCC CATCTAGTGT AGCGCTAAGT AGTCGTAAT CGTAAGCAAT	1140
TGTAAGGCAA ACACTTTCA AGTGAGCGAA ATATCAAGCA AACTGTGAGA ATTGAGGAC	1200
GTGTGACGAT GGAGCAACCC TTCCCCCCC GATCGAAAGA GAATATCATC AATCAACATT	1260
CCCGTGCCCG GAGGAGCTGC TCTTCATCA ACACCAACC CGAACTGGC CCTCAAAAGC	1320
CCGGCAACCT AAAGTTAGTC CTTTCATTAG CCTCTTCTAT CAATTAGTTA GTCAGCCAAC	1380
GTTTCTCTCT CTCTCATATAAT TCTAACCGAA AGTAAGCATA GAAAAGAACC AATACTTCAA	1440
TCAACATACC CACAAAAAAA AACAAATCCC CACCAACTGG CGCGGTACAA CACTGACCAA	1500
GGAGCG ATG GAC CAG CAA TTC TGC TTG CGC TGG AAC AAT CAT CCC ACA Met Asp Gln Gln Phe Cys Leu Arg Trp Asn Asn His Pro Thr	1548
1 5 10	
AAT TTG ACC GGC GTG CTA ACC TCA CTG CTG CAG CGG GAG GCG CTA TGC Asn Leu Thr Gly Val Leu Thr Ser Leu Leu Gln Arg Glu Ala Leu Cys	1596
15 20 25 30	
GAC GTC ACG CTC GCC TGC GAG GGC GAA ACA GTC AAG GCT CAC CAG ACC Asp Val Thr Leu Ala Cys Glu Gly Glu Thr Val Lys Ala His Gln Thr	1644
35 40 45	
ATC CTG TCA GCC TGC AGT CCG TAC TTC GAG ACG ATT TTC CTA CAG AAC Ile Leu Ser Ala Cys Ser Pro Tyr Phe Glu Thr Ile Phe Leu Gln Asn	1692
50 55 60	
CAG CAT CCA CAT CCC ATC ATC TAC TTG AAA GAT GTC AGA TAC TCA GAG Gln His Pro His Pro Ile Ile Tyr Leu Lys Asp Val Arg Tyr Ser Glu	1740
65 70 75	
ATG CGA TCT CTG CTC GAC TTC ATG TAC AAG GGC GAG GTC AAC GTG GGC Met Arg Ser Leu Leu Asp Phe Met Tyr Lys Gly Glu Val Asn Val Gly	1788
80 85 90	
CAG AGT TCG CTG CCC ATG TTT CTC AAG ACG GCC GAG AGC CTG CAG GTG Gln Ser Ser Leu Pro Met Phe Leu Lys Thr Ala Glu Ser Leu Gln Val	1836
95 100 105 110	
CGT GGT CTC ACA GAT AAC AAC AAT CTG AAC TAC CGC TCC GAC TGC GAC Arg Gly Leu Thr Asp Asn Asn Leu Asn Tyr Arg Ser Asp Cys Asp	1884
115 120 125	
AAG CTG CGC GAT TCG GCG AGT TCG CCG ACC GGA CGT GGG CCG AGT Lys Leu Arg Asp Ser Ala Ala Ser Ser Pro Thr Gly Arg Gly Pro Ser	1932
130 135 140	
AAT TAC ACT GGC GGC CTG GGC GGC GCT GGG GGC GTG GCC GAT GCG ATG Asn Tyr Thr Gly Gly Leu Gly Gly Ala Gly Gly Val Ala Asp Ala Met	1980
145 150 155	
CGC GAA TCC CGC GAC TCC CTG CGC TCC CGC TGC GAA CGG GAT CTG CGC Arg Glu Ser Arg Asp Ser Leu Arg Ser Arg Cys Glu Arg Asp Leu Arg	2028
160 165 170	
GAC GAG CTG ACG CAG CGC AGC AGC AGC AGC ATG AGC GAA CGC AGC TCG Asp Glu Leu Thr Gln Arg Ser Ser Ser Met Ser Glu Arg Ser Ser	2076
175 180 185 190	

GGC GCA GCA GCG GCG GCG GCG GCA GCA GCA GCG GTA GCG GCC GCC GGC Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Val Ala Ala Ala Gly 195 200 205	2124
GGC AAT GTG AAT GCG GCT GCC GTC GCC CTG GGC CTG ACC ACG CCC ACC Gly Asn Val Asn Ala Ala Val Ala Leu Gly Leu Thr Thr Pro Thr 210 215 220	2172
GCG GCG GCA GCT GCG GCG GTA GCA GCT GCG GTG GCA GCG GCC GCC AAT Ala Ala Ala Ala Ala Val Ala Ala Ala Val Ala Ala Ala Ala Asn 225 230 235	2220
CGA AGT GCC AGC GCC GAT GGA TGC AGC GAT CGG GGA AGC GAA CGC GGT Arg Ser Ala Ser Ala Asp Gly Cys Ser Asp Arg Gly Ser Glu Arg Gly 240 245 250	2268
ACG CTC GAG CGG ACG GAT AGT CGC GAT GAT CTA TTG CAG CTG GAT TAT Thr Leu Glu Arg Thr Asp Ser Arg Asp Asp Leu Leu Gln Leu Asp Tyr 255 260 265 270	2316
AGC AAC AAG GAT AAC AAC AAT AGC AAC AGC AGT AGT ACC ACC GGC GGC AAC Ser Asn Lys Asp Asn Asn Ser Asn Ser Ser Thr Gly Gly Asn 275 280 285	2364
AAC AAC AAC AAT AAT AAC AAC AAC AAT AGC AGC AGC AAC AAC AAC AAC Asn Asn Asn Asn Asn Asn Asn Ser Ser Ser Asn Asn Asn Asn Asn 290 295 300	2412
AAC AGC AGC AAT AGG GAG CGC AAC AAT AGC GGC GAA CGT GAG CGG Asn Ser Ser Asn Arg Glu Arg Asn Asn Ser Gly Glu Arg Glu Arg 305 310 315	2460
GAG CGA GAA AGA GAG CGT GAG CGG GAC AGG GAC AGG GAG CTG TCC ACC Glu Arg Glu Arg Glu Arg Glu Arg Asp Arg Asp Arg Glu Leu Ser Thr 320 325 330	2508
ACG CCG GTG GAG CAG CTG AGT AGT AAG CGC AGA CGT AAG AAC TCA Thr Pro Val Glu Gln Leu Ser Ser Lys Arg Arg Arg Lys Asn Ser 335 340 345 350	2556
TCA TCC AAC TGT GAT AAC TCG CTG TCC TCG AGC CAC CAG GAC AGG CAC Ser Ser Asn Cys Asp Asn Ser Leu Ser Ser His Gln Asp Arg His 355 360 365	2604
TAC CCG CAG GAC TCT CAG GCC AAC TTC AAG TCG AGT CCC GTG CCC AAA Tyr Pro Gln Asp Ser Gln Ala Asn Phe Lys Ser Ser Pro Val Pro Lys 370 375 380	2652
ACG GGC GGC AGC ACA TCG GAA TCG GAG GAC GCC GGC GGT CGC CAC GAC Thr Gly Ser Thr Ser Glu Ser Glu Asp Ala Gly Gly Arg His Asp 385 390 395	2700
TCG CCG CTG TCG ATG ACC ACA AGC GTT CAT CTG GGC GGC GGT GGT GGC Ser Pro Leu Ser Met Thr Ser Val His Leu Gly Gly Gly Gly 400 405 410	2748
AAT GTG GGC GCG GCC AGC GCC CTT AGC GGT CTG AGC CAG TCG CTG AGC Asn Val Gly Ala Ala Ser Ala Leu Ser Gly Leu Ser Gln Ser Leu Ser 415 420 425 430	2796
ATC AAG CAG GAG CTG ATG GAC GCC CAG CAG CAG CAG CAG CAT CGG GAA Ile Lys Gln Glu Leu Met Asp Ala Gln Gln Gln Gln His Arg Glu 435 440 445	2844
CAC CAC GTG GCC CTG CCC CCA GAT TAC TTG CCG AGC GCC GCT CTA AAG His His Val Ala Leu Pro Pro Asp Tyr Leu Pro Ser Ala Ala Leu Lys 450 455 460	2892

CTG CAC GCG GAG GAT ATG TCA ACG CTG CTC ACG CAG CAT GCT TTG CAA	2940
Leu His Ala Glu Asp Met Ser Thr Leu Leu Thr Gln His Ala Leu Gln	
465 470 475	
GCA GCA GAT GCG CGG GAC GAG CAC AAC GAC GCC AAA CAA CTG CAG CTG	2988
Ala Ala Asp Ala Arg Asp Glu His Asn Asp Ala Lys Gln Leu Gln Leu	
480 485 490	
GAC CAG ACG GAC AAT ATC GAC GGC AGC AGC GCC CGC CAC CAC CTG TCG	3036
Asp Gln Thr Asp Asn Ile Asp Gly Ser Ser Ala Arg His His Leu Ser	
495 500 505 510	
ACC CCC CTG TCG ACC TCG TCG GCC TCG CCC CCG CCG CCC CCT TTC	3084
Thr Pro Leu Ser Thr Ser Ser Ala Ser Pro Pro Pro Pro Pro Phe	
515 520 525	
GGG ATG CAC CTG TCG GCG GCC CTG AAA CGC GAG TAC CAT CCT CTG CAC	3132
Gly Met His Leu Ser Ala Ala Leu Lys Arg Glu Tyr His Pro Leu His	
530 535 540	
TAT ATG GCC GCC GGC AAC GGT CAC AAC GGC CCA TCG GCG CTT GGT TAT	3180
Tyr Met Ala Ala Gly Asn Gly His Asn Gly Pro Ser Ala Leu Gly Tyr	
545 550 555	
GGC AAT CAG GGA TCG GGC AAT GCG CCG AAT AGT GCC GGA GGA GCT GGA	3228
Gly Asn Gln Gly Ser Gly Asn Ala Pro Asn Ser Ala Gly Gly Ala Gly	
560 565 570	
TCG GTT GCG GGC GGA GTG GGA GCC GGC GGA GGA GCC GGC GGA GCA ACT	3276
Ser Val Ala Gly Gly Val Gly Ala Gly Gly Ala Gly Gly Ala Thr	
575 580 585 590	
GGA GCA GCT GGC CAT AAT TCG CAT CAC ACC ATG TCG TAC CAC AAC ATG	3324
Gly Ala Ala Gly His Asn Ser His His Thr Met Ser Tyr His Asn Met	
595 600 605	
TTC ACG CCG TCC CGC GAT CCG GGC ACC ATG TGG CGG TGC CGC TCC TGC	3372
Phe Thr Pro Ser Arg Asp Pro Gly Thr Met Trp Arg Cys Arg Ser Cys	
610 615 620	
GGC AAG GAG GTG ACC AAT CGC TGG CAC CAC TTT CAC TCC CAC ACC GCC	3420
Gly Lys Glu Val Thr Asn Arg Trp His His Phe His Ser His Thr Ala	
625 630 635	
CAG CGG TCC ATG TGT CCC TAC TGC CCG GCC ACC TAC AGC AGG ATC GAT	3468
Gln Arg Ser Met Cys Pro Tyr Cys Pro Ala Thr Tyr Ser Arg Ile Asp	
640 645 650	
ACG CTG CGC TCC CAT TTG CGG GTG AAG CAT CCG GAT CGC CTG CTC AAG	3516
Thr Leu Arg Ser His Leu Arg Val Lys His Pro Asp Arg Leu Leu Lys	
655 660 665 670	
CTG AAC TCG TCC ATT TAAGGGCGTG GCCGGGGCCC AAGTGCAGCC CATCACCGCC	3571
Leu Asn Ser Ser Ile	
675	
AGCTTTACCA GCAGCAACAA CAGCCGCATC ATAAGCAGAA GCAGAAGCAG CAACAGCAGC	3631
AGCAGCAACA GCAGCAGCAT CAGCCGCATC AGCAGCAACA GCAACCAGCT TACTACGTCA	3691
GCAACTATAG CAACTACAGC AATAATAGAT ACAGCTACAG CGATAGTTA TTGTAAATCG	3751
CTGCAGTTCT AGGTGGATT TTCTTGCATT TAGTCGTCGT CCAGTCGTGT ACATTACCCA	3811
CTAGCTATCC AAGCAATAAC CATAACCCAA ACTAGTAGAA AACCGAAGAT GCTATGCTAT	3871
GGCAAAACGT AAAGCGTTAA ACACAAAGTAT ATTGATAATC TTAACTAAC TTATTGATAAA	3931

ACTTTGACAC AATCGTCCCA TCAATTTATA AATGTGTATA ACTAAGGAAG ATTAGGAAAA	3991
GGTTTCAGTT GCGAGTCGAG GAGAAGGATA TGCCCAGCAT AGAGGGCCAG TGGAGGCAGA	4051
AAAAAAAGTTT TCCAAAGCCA CAACAAACCG TTTCGAAGGT TTCTAAATGT TGTTTCCTAA	4111
AAACTATAAA GTAATAACTA CACTAATACT AGAGAGAGAA AGTCGAGGAG AATCGTTTG	4171
AGCCGATTCA GCAAATTGGG GTCACTACCA CATCACCGG GGTCACCAGC AGCAGCAGCA	4231
GCAGCAGCAA ATGGAGGATG CGGATGCGAA TGCGGATGCG GATGAGGATC AGGATGAGGA	4291
TCAGCCAGCA CAGAACAGT CACCCACAAA TACTACTCAT ACGAAGGTCA CATTAGGTTT	4351
TAGTTTACTT TAATTGTAA TGTCTAGATT TTAGTGTAA CCGATATGTT CTGCGGAGTA	4411
GGAAACGGAT GAGGGCTACT CAACCAACTA CAAAGAAATT TTCATATACC TCAAATGCAT	4471
TTCAGTTTTA TTGTTGATTG CTTTAATTTC AGTCTACGTA GTCAGTTAGC ACTTATACAT	4531
AAAGTACAC ATACATATAT GTTATTTTT AATCGGTTCC AATTGAAATC GGCGAGATAG	4591
CCAATAGTTT ACCAATGTTT TCCTCTGTTT TTTAGTGTGT GTGGTGTGTT CCCTATCACT	4651
ATCACACTTT TGATTTGTC CTATGCGTTA AGTTGAAGAT TTTAGGATTA GCTCGAACCA	4711
CTTGAACAC ACCAACACAC TTTGTTAAGC TTGTTATAT TTTATATTAA TGTCACACG	4771
TTTATTTAGT TAAAGTACAC TAAACACATA TGAAATCACG CGGAAGAAAG TTAGTTGATA	4831
TGAG	4835

## (2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 675 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met	Asp	Gln	Gln	Phe	Cys	Leu	Arg	Trp	Asn	Asn	His	Pro	Thr	Asn	Leu
1						5			10						15
Thr	Gly	Val	Leu	Thr	Ser	Leu	Leu	Gln	Arg	Glu	Ala	Leu	Cys	Asp	Val
								20		25					30
Thr	Leu	Ala	Cys	Glu	Gly	Glu	Thr	Val	Lys	Ala	His	Gln	Thr	Ile	Leu
							35		40						45
Ser	Ala	Cys	Ser	Pro	Tyr	Phe	Glu	Thr	Ile	Phe	Leu	Gln	Asn	Gln	His
						50			55			60			
Pro	His	Pro	Ile	Ile	Tyr	Leu	Lys	Asp	Val	Arg	Tyr	Ser	Glu	Met	Arg
						65			70		75				80
Ser	Leu	Leu	Asp	Phe	Met	Tyr	Lys	Gly	Glu	Val	Asn	Val	Gly	Gln	Ser
							85			90					95
Ser	Leu	Pro	Met	Phe	Leu	Lys	Thr	Ala	Glu	Ser	Leu	Gln	Val	Arg	Gly
							100			105					110
Leu	Thr	Asp	Asn	Asn	Leu	Asn	Tyr	Arg	Ser	Asp	Cys	Asp	Lys	Leu	
							115			120					125

Arg Asp Ser Ala Ala Ser Ser Pro Thr Gly Arg Gly Pro Ser Asn Tyr  
 130 135 140  
 Thr Gly Gly Leu Gly Gly Ala Gly Gly Val Ala Asp Ala Met Arg Glu  
 145 150 155 160  
 Ser Arg Asp Ser Leu Arg Ser Arg Cys Glu Arg Asp Leu Arg Asp Glu  
 165 170 175  
 Leu Thr Gln Arg Ser Ser Ser Met Ser Glu Arg Ser Ser Ala Ala  
 180 185 190  
 Ala Ala Ala Ala Ala Ala Ala Val Ala Ala Ala Gly Gly Asn  
 195 200 205  
 Val Asn Ala Ala Ala Val Ala Leu Gly Leu Thr Thr Pro Thr Ala Ala  
 210 215 220  
 Ala Ala Ala Ala Val Ala Ala Val Ala Ala Ala Asn Arg Ser  
 225 230 235 240  
 Ala Ser Ala Asp Gly Cys Ser Asp Arg Gly Ser Glu Arg Gly Thr Leu  
 245 250 255  
 Glu Arg Thr Asp Ser Arg Asp Asp Leu Leu Gln Leu Asp Tyr Ser Asn  
 260 265 270  
 Lys Asp Asn Asn Asn Ser Asn Ser Ser Ser Thr Gly Gly Asn Asn Asn  
 275 280 285  
 Asn Asn Asn Asn Asn Asn Asn Ser Ser Ser Asn Asn Asn Asn Ser  
 290 295 300  
 Ser Ser Asn Arg Glu Arg Asn Asn Ser Gly Glu Arg Glu Arg Glu Arg  
 305 310 315 320  
 Glu Arg Glu Arg Glu Arg Asp Arg Asp Arg Glu Leu Ser Thr Thr Pro  
 325 330 335  
 Val Glu Gln Leu Ser Ser Lys Arg Arg Arg Lys Asn Ser Ser Ser  
 340 345 350  
 Asn Cys Asp Asn Ser Leu Ser Ser Ser His Gln Asp Arg His Tyr Pro  
 355 360 365  
 Gln Asp Ser Gln Ala Asn Phe Lys Ser Ser Pro Val Pro Lys Thr Gly  
 370 375 380  
 Gly Ser Thr Ser Glu Ser Glu Asp Ala Gly Gly Arg His Asp Ser Pro  
 385 390 395 400  
 Leu Ser Met Thr Thr Ser Val His Leu Gly Gly Gly Gly Asn Val  
 405 410 415  
 Gly Ala Ala Ser Ala Leu Ser Gly Leu Ser Gln Ser Leu Ser Ile Lys  
 420 425 430  
 Gln Glu Leu Met Asp Ala Gln Gln Gln Gln His Arg Glu His His  
 435 440 445  
 Val Ala Leu Pro Pro Asp Tyr Leu Pro Ser Ala Ala Leu Lys Leu His  
 450 455 460  
 Ala Glu Asp Met Ser Thr Leu Leu Thr Gln His Ala Leu Gln Ala Ala  
 465 470 475 480

Asp Ala Arg Asp Glu His Asn Asp Ala Lys Gln Leu Gln Leu Asp Gln  
 485 490 495  
 Thr Asp Asn Ile Asp Gly Ser Ser Ala Arg His His Leu Ser Thr Pro  
 500 505 510  
 Leu Ser Thr Ser Ser Ala Ser Pro Pro Pro Pro Phe Gly Met  
 515 520 525  
 His Leu Ser Ala Ala Leu Lys Arg Glu Tyr His Pro Leu His Tyr Met  
 530 535 540  
 Ala Ala Gly Asn Gly His Asn Gly Pro Ser Ala Leu Gly Tyr Gly Asn  
 545 550 555 560  
 Gln Gly Ser Gly Asn Ala Pro Asn Ser Ala Gly Gly Ala Gly Ser Val  
 565 570 575  
 Ala Gly Gly Val Gly Ala Gly Gly Ala Gly Gly Ala Thr Gly Ala  
 580 585 590  
 Ala Gly His Asn Ser His His Thr Met Ser Tyr His Asn Met Phe Thr  
 595 600 605  
 Pro Ser Arg Asp Pro Gly Thr Met Trp Arg Cys Arg Ser Cys Gly Lys  
 610 615 620  
 Glu Val Thr Asn Arg Trp His His Phe His Ser His Thr Ala Gln Arg  
 625 630 635 640  
 Ser Met Cys Pro Tyr Cys Pro Ala Thr Tyr Ser Arg Ile Asp Thr Leu  
 645 650 655  
 Arg Ser His Leu Arg Val Lys His Pro Asp Arg Leu Leu Lys Leu Asn  
 660 665 670  
 Ser Ser Ile  
 675

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 608 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: EcoRI genomic clone  
containing 3 dsx repeats
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 324..420
  - (D) OTHER INFORMATION: /note= "where N has not  
been precisely determined"
- (ix) FEATURE:
  - (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 483..485
  - (D) OTHER INFORMATION: /note= "where N has not  
been precisely determined"

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 509..509
- (D) OTHER INFORMATION: /note= "where N has not been precisely determined"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAATTCGAGG ACGTGTGACG ATGGAGCAAC CCTTCCCCCC CAGATCGAAA GAGAATATCA	60
TCAATCAACA TTCCCGTGCC CGGAGGAGCG GCTCTTCAAT CAACACTCAA CCCGAACCTGG	120
GCCCTCAAAA GCCCGGCAAC CTAAAGTTAG TCTTCATTA GCCTCTTCTA TCAATTAGGT	180
AGTCAGCCAA CGTTTCTCTC TCTCTCATAA TTCTAACCGA AAGTAAGCAT AGAAAAGAAC	240
CAATACTTCA ATCAACATAC CCACAAAAAA AAACAAATCC CCACCAACTG GCGTCGGTAA	300
GTGAAGAGCC ATTTTAATTA TAGNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN	360
NNNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN	420
TGATCGCCGA TGATGCATGT GATAAGCAAG TGATGAACAA TCCGTAGCAA TCAGGCAGTA	480
GGNNNCTTGA ACAAAATTAA CTTAGCTGNA TTTTGCATGCCAAATGAA AAATAACAAA	540
CCGTAAATTC CAATGGTAAC TAAAACTAGC AATACTAACT CTAGCCGATG GAACATGCAA	600
CCGAATTC	608

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1244 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: cDNA to mRNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: alternative 3' end  
starting at nt. 3012 of SEQ ID NO:9

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..1021

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

T CGC GTC AAG TGT TTT AAC ATT AAG CAC GAC CGT CAT CCG GAT CGG	46
Arg Val Lys Cys Phe Asn Ile Lys His Asp Arg His Pro Asp Arg	
1 5 10 15	
GAA CTG GAT CGA AAT CAT CGG GAG CAC GAC GAC GAT CCA GGC GTT ATC	94
Glu Leu Asp Arg Asn His Arg Glu His Asp Asp Asp Pro Gly Val Ile	
20 25 30	
GAG GAG GTC GTT GTG GAT CAC GTT CGT GAG ATG GAA GCG GGG AAT GAG	142
Glu Glu Val Val Asp His Val Arg Glu Met Glu Ala Gly Asn Glu	
35 40 45	

CAC GAT CCG GAG GAG ATG AAG GAG GCA GCC TAC CAT GCC ACA CCG CCC	190
His Asp Pro Glu Glu Met Lys Glu Ala Ala Tyr His Ala Thr Pro Pro	
50 55 60	
AAG TAC AGA CGG GCT GTG TAT GCT CCT CCG CAT CCG GAT GAA GAG	238
Lys Tyr Arg Arg Ala Val Val Tyr Ala Pro Pro His Pro Asp Glu Glu	
65 70 75	
GCG GCC TCC GGA TCG GGA TCG GAT ATC TAT GTG GAT GGC GGC TAC AAT	286
Ala Ala Ser Gly Ser Gly Ser Asp Ile Tyr Val Asp Gly Gly Tyr Asn	
80 85 90 95	
TGC GAG TAC AAG TGC AAG GAG CTC AAC ATG CAG CGC AAC ATA CGA TGC	334
Cys Glu Tyr Lys Cys Lys Glu Leu Asn Met Gln Arg Asn Ile Arg Cys	
100 105 110	
AGT CGC CAG CAG CAC ATG ATG TCC CAC TAT TCG CCG CAT CAT CCG CAC	382
Ser Arg Gln Gln His Met Ser His Tyr Ser Pro His His Pro His	
115 120 125	
CAT CGA TCC CTC ATA GAT TGC CCC GCC GAG GCG GCT TAC TCA CCG CCG	430
His Arg Ser Leu Ile Asp Cys Pro Ala Glu Ala Ala Tyr Ser Pro Pro	
130 135 140	
G TG GCC AAC AAT CAG GCC TAC CTG GCC AGC AAT GGA GCG GTG CAG CAG	478
Val Ala Asn Asn Gln Ala Tyr Leu Ala Ser Asn Gly Ala Val Gln Gln	
145 150 155	
TTG GAT TTG AGC ACT TAC CAT GGC CAC GCA AAC CAC CAA CTC CAC CAG	526
Leu Asp Leu Ser Thr Tyr His Gly His Ala Asn His Gln Leu His Gln	
160 165 170 175	
CAT CCG CCA TCA GCC ACA CAT CCC AGT CAC TCG CAG AGC TCA CCC CAT	574
His Pro Pro Ser Ala Thr His Pro Ser His Ser Gln Ser Ser Pro His	
180 185 190	
TAT CCA AGC GCC TCT GGT GCA GGT GCT GGC GCG GGT TCA GTC TCG GTT	622
Tyr Pro Ser Ala Ser Gly Ala Gly Ala Gly Ser Val Ser Val	
195 200 205	
TCA ATA GCA GGA TCT GCA TCG GGA TCA GCC ACA TCT GCA CCA GCT TCG	670
Ser Ile Ala Gly Ser Ala Ser Gly Ser Ala Thr Ser Ala Pro Ala Ser	
210 215 220	
GTG GCC ACG TCA GCG GTC TCG CCG CAG CCG AGC TCC AGT TCC ACT GGA	718
Val Ala Thr Ser Ala Val Ser Pro Gln Pro Ser Ser Ser Thr Gly	
225 230 235	
TCC ACA TCG TCG GCG GCG GCG GTT GCA GCG GCA GCT GCT GCG GCT GCC	766
Ser Thr Ser Ser Ala Ala Ala Val Ala Ala Ala Ala Ala Ala Ala Ala	
240 245 250 255	
AAT CGG CGG GAT CAC AAC ATT GAC TAC TCC ACC CTG TTT GTC CAG CTA	814
Asn Arg Arg Asp His Asn Ile Asp Tyr Ser Thr Leu Phe Val Gln Leu	
260 265 270	
TCG GGC ACG TTG CCC ACT CTA TAC CGA TGC GTT AGT TGC AAC AAG ATC	862
Ser Gly Thr Leu Pro Thr Leu Tyr Arg Cys Val Ser Cys Asn Lys Ile	
275 280 285	
GTG TCC AAT CGC TGG CAC CAT GCC AAT ATC CAT CGA CCG CAG AGT CAT	910
Val Ser Asn Arg Trp His His Ala Asn Ile His Arg Pro Gln Ser His	
290 295 300	
GAG TGC CCC GTT TGC GGG CAG AAA TTC ACT CGC AGG GAC AAT ATG AAG	958
Glu Cys Pro Val Cys Gly Gln Lys Phe Thr Arg Arg Asp Asn Met Lys	
305 310 315	

GCG CAC TGT AAG ATC AAG CAT GCG GAC ATC AAG GAT CGA TTC TTT AGC	1006
Ala His Cys Lys Ile Lys His Ala Asp Ile Lys Asp Arg Phe Phe Ser	
320 325 330 335	
CAC TAT GTA CAT ATG TGATCACTTC TCTAGGCAGG CAGCAAAACA AATCAAATCA	1061
His Tyr Val His Met	
340	
AAAAATCAGT AACAGATCGA ATGGTTTCAGCTAAGTA ACCAAGAACGTA AAGCAAACGT	1121
ATACGTAATC CAGAGTGAGG AGCCAACAGC CATCAGTTGG ATGTACATCT ATATCTATAT	1181
CTATACATTT ATAAACCCCTA TCAGAAAACA GACTCGTGCC GAATTCAATAT CAAGCTTATC	1241
CAT	1244

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 340 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Arg Val Lys Cys Phe Asn Ile Lys His Asp Arg His Pro Asp Arg Glu	
1 5 10 15	
Leu Asp Arg Asn His Arg Glu His Asp Asp Asp Pro Gly Val Ile Glu	
20 25 30	
Glu Val Val Val Asp His Val Arg Glu Met Glu Ala Gly Asn Glu His	
35 40 45	
Asp Pro Glu Glu Met Lys Glu Ala Ala Tyr His Ala Thr Pro Pro Lys	
50 55 60	
Tyr Arg Arg Ala Val Val Tyr Ala Pro Pro His Pro Asp Glu Glu Ala	
65 70 75 80	
Ala Ser Gly Ser Gly Ser Asp Ile Tyr Val Asp Gly Gly Tyr Asn Cys	
85 90 95	
Glu Tyr Lys Cys Lys Glu Leu Asn Met Gln Arg Asn Ile Arg Cys Ser	
100 105 110	
Arg Gln Gln His Met Met Ser His Tyr Ser Pro His His Pro His His	
115 120 125	
Arg Ser Leu Ile Asp Cys Pro Ala Glu Ala Ala Tyr Ser Pro Pro Val	
130 135 140	
Ala Asn Asn Gln Ala Tyr Leu Ala Ser Asn Gly Ala Val Gln Gln Leu	
145 150 155 160	
Asp Leu Ser Thr Tyr His Gly His Ala Asn His Gln Leu His Gln His	
165 170 175	
Pro Pro Ser Ala Thr His Pro Ser His Ser Gln Ser Ser Pro His Tyr	
180 185 190	
Pro Ser Ala Ser Gly Ala Gly Ala Gly Ser Val Ser Val Ser	
195 200 205	

Ile Ala Gly Ser Ala Ser Gly Ser Ala Thr Ser Ala Pro Ala Ser Val  
 210 215 220  
 Ala Thr Ser Ala Val Ser Pro Gln Pro Ser Ser Ser Ser Thr Gly Ser  
 225 230 235 240  
 Thr Ser Ser Ala Ala Ala Val Ala Ala Ala Ala Ala Ala Asn  
 245 250 255  
 Arg Arg Asp His Asn Ile Asp Tyr Ser Thr Leu Phe Val Gln Leu Ser  
 260 265 270  
 Gly Thr Leu Pro Thr Leu Tyr Arg Cys Val Ser Cys Asn Lys Ile Val  
 275 280 285  
 Ser Asn Arg Trp His His Ala Asn Ile His Arg Pro Gln Ser His Glu  
 290 295 300  
 Cys Pro Val Cys Gly Gln Lys Phe Thr Arg Arg Asp Asn Met Lys Ala  
 305 310 315 320  
 His Cys Lys Ile Lys His Ala Asp Ile Lys Asp Arg Phe Phe Ser His  
 325 330 335  
 Tyr Val His Met  
 340

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4255 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (C) INDIVIDUAL ISOLATE: fruitless transcript in Fig. 7E
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1507..4032

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAATTGGCA CGAGATTAC	CTATGGCATA TCATCAGCAA	CACACATCAA CGCAC	TTCTC	60
TGCTATGTCT GCAATCAACC	AAAATATCAA AAAAAAAAAG	AAAACAAAAA AGAGTCAACA		120
TCAATTAA AGTTTTACG	TTGGTTCGAA AGAGTTAAA	ATGCCCTTAA CTATTAACGC		180
CCAAAAGTAA ACGTAGATTA	AAGTAATATT AGCCAATCAA	TCGTAAAATA TCAGCTTCG		240
TTTTTAAAGA CTTACCAATG	GACTTTGATC CCATCAATTG	CAAATCTAAA GTAGAGAAAT		300
AGAGAGAGAT AAGAGATATA	ATATCACTAA CCAAAAGTGT	TTGCCACGAG TATTAAAATG		360
TTAACTACTA CAATAGAATA	CGTATTCTTG TTTCCCTCGC	TAGTATGTAT AAGCAAAC		420
ACTGCAAGAA ACAACACCAA	CTAATTAATA TTTAATAGCA	TAATGGTAAT ATCGTAAGAA		480

TATCATAGAT TTAAGGCAGA GCATTTAGA CAGCACTTGT ACCGTTCTAG ACTTAAGTAT	540
TCGAAGTATA CGTAACTCAA GCAATCCAAT AACAAATAACT AAGTAGAAGT TCTTTTCAAA	600
ATAATACTAT ACACGAATCC TTCAGTCAAA CCCCTACAA TATTACTTAG ATAAACATAT	660
AGTATTATAT AGCCAAAGCC AGGAAAGGAG TTGTAAGCCA TTGCATATAT ATATTTGGTA	720
GATAAAGAAC AGCTAACGAA AGGGTCCACA AGCTACCCAT AACTTACTTA GAATAACTAA	780
ACACAACTAG CCAAGAAGTA GATATCTATA TATATATCGA GTTTGCTAA CATCAAAGTA	840
TACGTAAATT GAAAACCAAG AATTTGCCT AGCTTAAATA ACACTCTTC AAAGCAATAC	900
CATAAAACAAT AATTACAAGT TAACGCAACT AAACACATAT TGTATACCAG ATAGTTTATG	960
CCTAAACACT ACTAGTAGCC CTAAGTCCTA GGCATAAACC GAGCACCACG GCGAGATATG	1020
CACCCATGTA AAATGCAGAA ATTAATTACC AAGAGTACAA ACTGTAAAGG AAACCCCTAT	1080
TGAAGCTCAA TTGGCCAGCC CATCTAGTGT AGCGCTAAGT AGTCGTAAT CGTAAGCAAT	1140
TGTAAGGCCA ACACTTTCAGA AGTGGCGAA ATATCAAGCA AACTGTGAGA ATTCGAGGAC	1200
GTGTGACGAT GGAGCAACCC TTCCCCCCC GATCGAAAGA GAATATCATC AATCAACATT	1260
CCCGTGCCTG GAGGAGCTGC TCTTCATCA ACACCAACC CGAACTGGGC CCTCAAAAGC	1320
CCGGCAACCT AAAGTTAGTC CTTTCATTAG CCTCTTCTAT CAATTAGTTA GTCAGCCAAC	1380
GTTTCTCTCT CTCTCATAAT TCTAACCGAA AGTAAGCATA GAAAAGAACC AATACTTCAA	1440
TCAACATACC CACAAAAAAA AACAAATCCC CACCAACTGG CGCGGTACAA CACTGACCAA	1500
GGAGCG ATG GAC CAG CAA TTC TGC TTG CGC TGG AAC AAT CAT CCC ACA	1548
Met Asp Gln Gln Phe Cys Leu Arg Trp Asn Asn His Pro Thr	
1 5 10	
AAT TTG ACC GGC GTG CTA ACC TCA CTG CTG CAG CGG GAG GCG CTA TGC	1596
Asn Leu Thr Gly Val Leu Thr Ser Leu Leu Gln Arg Glu Ala Leu Cys	
15 20 25 30	
GAC GTC ACG CTC GCC TGC GAG GGC GAA ACA GTC AAG GCT CAC CAG ACC	1644
Asp Val Thr Leu Ala Cys Glu Gly Glu Thr Val Lys Ala His Gln Thr	
35 40 45	
ATC CTG TCA GCC TGC AGT CCG TAC TTC GAG ACG ATT TTC CTA CAG AAC	1692
Ile Leu Ser Ala Cys Ser Pro Tyr Phe Glu Thr Ile Phe Leu Gln Asn	
50 55 60	
CAG CAT CCA CAT CCC ATC ATC TAC TTG AAA GAT GTC AGA TAC TCA GAG	1740
Gln His Pro His Pro Ile Ile Tyr Leu Lys Asp Val Arg Tyr Ser Glu	
65 70 75	
ATG CGA TCT CTG CTC GAC TTC ATG TAC AAG GGC GAG GTC AAC GTG GGC	1788
Met Arg Ser Leu Leu Asp Phe Met Tyr Lys Gly Glu Val Asn Val Gly	
80 85 90	
CAG AGT TCG CTG CCC ATG TTT CTC AAG ACG GCC GAG AGC CTG CAG GTG	1836
Gln Ser Ser Leu Pro Met Phe Leu Lys Thr Ala Glu Ser Leu Gln Val	
95 100 105 110	
CGT GGT CTC ACA GAT AAC AAC AAT CTG AAC TAC CGC TCC GAC TGC GAC	1884
Arg Gly Leu Thr Asp Asn Asn Leu Asn Tyr Arg Ser Asp Cys Asp	
115 120 125	

AAG CTG CGC GAT TCG GCG GCC AGT TCG CCG ACC GGA CGT GGG CCG AGT Lys Leu Arg Asp Ser Ala Ala Ser Ser Pro Thr Gly Arg Gly Pro Ser 130 135 140	1932
AAT TAC ACT GGC GGC CTG GGC GCT GGG GGC GTG GCC GAT GCG ATG Asn Tyr Thr Gly Gly Leu Gly Ala Gly Gly Val Ala Asp Ala Met 145 150 155	1980
CGC GAA TCC CGC GAC TCC CTG CGC TCC CGC TGC GAA CGG GAT CTG CGC Arg Glu Ser Arg Asp Ser Leu Arg Ser Arg Cys Glu Arg Asp Leu Arg 160 165 170	2028
GAC GAG CTG ACG CAG CGC AGC AGC AGC ATG AGC GAA CGC AGC TCG Asp Glu Leu Thr Gln Arg Ser Ser Ser Met Ser Glu Arg Ser Ser 175 180 185 190	2076
GCG GCA GCA GCG GCG GCG GCA GCA GCA GCG GTA GCG GCC GCC GGC Ala Ala Ala Ala Ala Ala Ala Ala Val Ala Ala Ala Gly 195 200 205	2124
GGC AAT GTG AAT GCG GCT GCC GTC GCC CTG GGC CTG ACC ACG CCC ACC Gly Asn Val Asn Ala Ala Val Ala Leu Gly Leu Thr Thr Pro Thr 210 215 220	2172
GCG GCG GCA GCT GCG GCG GTA GCA GCT GCG GTG GCA GCG GCC GCC AAT Ala Ala Ala Ala Ala Val Ala Ala Val Ala Ala Ala Ala Asn 225 230 235	2220
CGA AGT GCC AGC GCC GAT GGA TGC AGC GAT CGG GGA AGC GAA CGC GGT Arg Ser Ala Ser Ala Asp Gly Cys Ser Asp Arg Gly Ser Glu Arg Gly 240 245 250	2268
ACG CTC GAG CGG ACG GAT AGT CGC GAT GAT CTA TTG CAG CTG GAT TAT Thr Leu Glu Arg Thr Asp Ser Arg Asp Asp Leu Leu Gln Leu Asp Tyr 255 260 265 270	2316
AGC AAC AAG GAT AAC AAC AAT AGC AAC AGC AGT AGT ACC ACC GGC GGC AAC Ser Asn Lys Asp Asn Asn Ser Asn Ser Ser Thr Gly Gly Asn 275 280 285	2364
AAC AAC AAC AAT AAT AAC AAC AAT AGC AGC AGC AAC AAC AAC AAC Asn Asn Asn Asn Asn Asn Asn Ser Ser Ser Asn Asn Asn Asn 290 295 300	2412
AAC AGC AGC AAT AGG GAG CGC AAC AAT AGC GGC GAA CGT GAG CGG Asn Ser Ser Asn Arg Glu Arg Asn Asn Ser Gly Glu Arg Glu Arg 305 310 315	2460
GAG CGA GAA AGA GAG CGT GAG CGG GAC AGG GAC AGG GAG CTG TCC ACC Glu Arg Glu Arg Glu Arg Glu Arg Asp Arg Asp Arg Glu Leu Ser Thr 320 325 330	2508
ACG CCG GTG GAG CAG CTG AGT AGT AAG CGC AGA CGT AAG AAC TCA Thr Pro Val Glu Gln Leu Ser Ser Lys Arg Arg Arg Lys Asn Ser 335 340 345 350	2556
TCA TCC AAC TGT GAT AAC TCG CTG TCC TCG AGC CAC CAG GAC AGG CAC Ser Ser Asn Cys Asp Asn Ser Leu Ser Ser His Gln Asp Arg His 355 360 365	2604
TAC CCG CAG GAC TCT CAG GCC AAC TTC AAG TCG AGT CCC GTG CCC AAA Tyr Pro Gln Asp Ser Gln Ala Asn Phe Lys Ser Ser Pro Val Pro Lys 370 375 380	2652
ACG GGC GGC AGC ACA TCG GAA TCG GAG GAC GCC GGC GGT CGC CAC GAC Thr Gly Ser Thr Ser Glu Ser Glu Asp Ala Gly Gly Arg His Asp 385 390 395	2700

TCG CCG CTG TCG ATG ACC ACA AGC GTT CAT CTG GGC GGC GGT GGT GGC Ser Pro Leu Ser Met Thr Thr Ser Val His Leu Gly Gly Gly Gly Gly 400 405 410	2748
AAT GTG GGC GCG GCC AGC GCC CTT AGC GGT CTG AGC CAG TCG CTG AGC Asn Val Gly Ala Ala Ser Ala Leu Ser Gly Leu Ser Gln Ser Leu Ser 415 420 425 430	2796
ATC AAG CAG GAG CTG ATG GAC GCC CAG CAG CAG CAG CAT CGG GAA Ile Lys Gln Glu Leu Met Asp Ala Gln Gln Gln Gln His Arg Glu 435 440 445	2844
CAC CAC GTG GCC CTG CCC CCA GAT TAC TTG CCG AGC GCC GCT CTA AAG His His Val Ala Leu Pro Pro Asp Tyr Leu Pro Ser Ala Ala Leu Lys 450 455 460	2892
CTG CAC GCG GAG GAT ATG TCA ACG CTG CTC ACG CAG CAT GCT TTG CAA Leu His Ala Glu Asp Met Ser Thr Leu Leu Thr Gln His Ala Leu Gln 465 470 475	2940
GCA GCA GAT GCG CGG GAC GAG CAC AAC GAC GCC AAA CAA CTG CAG CTG Ala Ala Asp Ala Arg Asp Glu His Asn Asp Ala Lys Gln Leu Gln Leu 480 485 490	2988
GAC CAG ACG GAC AAT ATC GAC GGT CGC GTC AAG TGT TTT AAC ATT AAG Asp Gln Thr Asp Asn Ile Asp Gly Arg Val Lys Cys Phe Asn Ile Lys 495 500 505 510	3036
CAC GAC CGT CAT CCG GAT CGG GAA CTG GAT CGA AAT CAT CGG GAG CAC His Asp Arg His Pro Asp Arg Glu Leu Asp Arg Asn His Arg Glu His 515 520 525	3084
GAC GAC GAT CCA GGC GTT ATC GAG GAG GTC GTT GTG GAT CAC GTT CGT Asp Asp Asp Pro Gly Val Ile Glu Glu Val Val Val Asp His Val Arg 530 535 540	3132
GAG ATG GAA GCG GGG AAT GAG CAC GAT CCG GAG GAG ATG AAG GAG GCA Glu Met Glu Ala Gly Asn Glu His Asp Pro Glu Glu Met Lys Glu Ala 545 550 555	3180
GCC TAC CAT GCC ACA CCG CCC AAG TAC AGA CGG GCT GTG GTT TAT GCT Ala Tyr His Ala Thr Pro Pro Lys Tyr Arg Arg Ala Val Val Tyr Ala 560 565 570	3228
CCT CCG CAT CCG GAT GAA GAG GCG GGC TCC GGA TCG GGA TCG GAT ATC Pro Pro His Pro Asp Glu Glu Ala Ala Ser Gly Ser Gly Ser Asp Ile 575 580 585 590	3276
TAT GTG GAT GGC GGC TAC AAT TGC GAG TAC AAG TGC AAG GAG CTC AAC Tyr Val Asp Gly Gly Tyr Asn Cys Glu Tyr Lys Cys Lys Glu Leu Asn 595 600 605	3324
ATG CAG CGC AAC ATA CGA TGC AGT CGC CAG CAG CAC ATG ATG TCC CAC Met Gln Arg Asn Ile Arg Cys Ser Arg Gln Gln His Met Met Ser His 610 615 620	3372
TAT TCG CCG CAT CAT CCG CAC CAT CGA TCC CTC ATA GAT TGC CCC GCC Tyr Ser Pro His His Pro His His Arg Ser Leu Ile Asp Cys Pro Ala 625 630 635	3420
GAG GCG GCT TAC TCA CCG CCG GTG GCC AAC AAT CAG GCC TAC CTG GCC Glu Ala Ala Tyr Ser Pro Pro Val Ala Asn Asn Gln Ala Tyr Leu Ala 640 645 650	3468
AGC AAT GGA GCG GTG CAG CAG TTG GAT TTG AGC ACT TAC CAT GGC CAC Ser Asn Gly Ala Val Gln Gln Leu Asp Leu Ser Thr Tyr His Gly His 655 660 665 670	3516

GCA AAC CAC CAA CTC CAC CAG CAT CCG CCA TCA GCC ACA CAT CCC AGT Ala Asn His Gln Leu His Gln His Pro Pro Ser Ala Thr His Pro Ser 675 680 685	3564
CAC TCG CAG AGC TCA CCC CAT TAT CCA AGC GCC TCT GGT GCA GGT GCT His Ser Gln Ser Ser Pro His Tyr Pro Ser Ala Ser Gly Ala Gly Ala 690 695 700	3612
GGC GCG GGT TCA GTC TCG GTT TCA ATA GCA GGA TCT GCA TCG GGA TCA Gly Ala Gly Ser Val Ser Val Ser Ile Ala Gly Ser Ala Ser Gly Ser 705 710 715	3660
GCC ACA TCT GCA CCA GCT TCG GTG GCC ACG TCA GCG GTC TCG CCG CAG Ala Thr Ser Ala Pro Ala Ser Val Ala Thr Ser Ala Val Ser Pro Gln 720 725 730	3708
CCG AGC TCC AGT TCC ACT GGA TCC ACA TCG TCG GCG GCG GCG GTT GCA Pro Ser Ser Ser Ser Thr Gly Ser Thr Ser Ser Ala Ala Ala Val Ala 735 740 745 750	3756
GCG GCA GCT GCT GCG GCT GCC AAT CGG CGG GAT CAC AAC ATT GAC TAC Ala Ala Ala Ala Ala Asn Arg Arg Asp His Asn Ile Asp Tyr 755 760 765	3804
TCC ACC CTG TTT GTC CAG CTA TCG GGC ACG TTG CCC ACT CTA TAC CGA Ser Thr Leu Phe Val Gln Leu Ser Gly Thr Leu Pro Thr Leu Tyr Arg 770 775 780	3852
TGC GTT AGT TGC AAC AAG ATC GTG TCC AAT CGC TGG CAC CAT GCC AAT Cys Val Ser Cys Asn Lys Ile Val Ser Asn Arg Trp His His Ala Asn 785 790 795	3900
ATC CAT CGA CCG CAG AGT CAT GAG TGC CCC GTT TGC GGG CAG AAA TTC Ile His Arg Pro Gln Ser His Glu Cys Pro Val Cys Gly Gln Lys Phe 800 805 810	3948
ACT CGC AGG GAC AAT ATG AAG GCG CAC TGT AAG ATC AAG CAT GCG GAC Thr Arg Arg Asp Asn Met Lys Ala His Cys Lys Ile Lys His Ala Asp 815 820 825 830	3996
ATC AAG GAT CGA TTC TTT AGC CAC TAT GTA CAT ATG TGATCACTTC Ile Lys Asp Arg Phe Phe Ser His Tyr Val His Met 835 840	4042
TCTAGGCAGG CAGCAAAACA AATCAAATCA AAAAATCAGT AACAGATCGA ATGGTTTCA CAGCTAAGTA ACCAAGAACATC AAGCAAACGT ATACGTAATC CAGAGTGAGG AGCCAACAGC CATCAGTTGG ATGTACATCT ATATCTATAT CTATACATTT ATAAACCCCTA TCAGAAAACA GACTCGTGCC GAATTCAATAT CAAGCTTATC CAT	4102 4162 4222 4255

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 842 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Asp Gln Gln Phe Cys Leu Arg Trp Asn Asn His Pro Thr Asn Leu	
1 5 10 15	

Thr Gly Val Leu Thr Ser Leu Leu Gln Arg Glu Ala Leu Cys Asp Val  
 20 25 30

Thr Leu Ala Cys Glu Gly Glu Thr Val Lys Ala His Gln Thr Ile Leu  
 35 40 45

Ser Ala Cys Ser Pro Tyr Phe Glu Thr Ile Phe Leu Gln Asn Gln His  
 50 55 60

Pro His Pro Ile Ile Tyr Leu Lys Asp Val Arg Tyr Ser Glu Met Arg  
 65 70 75 80

Ser Leu Leu Asp Phe Met Tyr Lys Gly Glu Val Asn Val Gly Gln Ser  
 85 90 95

Ser Leu Pro Met Phe Leu Lys Thr Ala Glu Ser Leu Gln Val Arg Gly  
 100 105 110

Leu Thr Asp Asn Asn Leu Asn Tyr Arg Ser Asp Cys Asp Lys Leu  
 115 120 125

Arg Asp Ser Ala Ala Ser Ser Pro Thr Gly Arg Gly Pro Ser Asn Tyr  
 130 135 140

Thr Gly Gly Leu Gly Gly Ala Gly Gly Val Ala Asp Ala Met Arg Glu  
 145 150 155 160

Ser Arg Asp Ser Leu Arg Ser Arg Cys Glu Arg Asp Leu Arg Asp Glu  
 165 170 175

Leu Thr Gln Arg Ser Ser Ser Met Ser Glu Arg Ser Ser Ala Ala  
 180 185 190

Ala Ala Ala Ala Ala Ala Ala Val Ala Ala Ala Gly Gly Asn  
 195 200 205

Val Asn Ala Ala Ala Val Ala Leu Gly Leu Thr Thr Pro Thr Ala Ala  
 210 215 220

Ala Ala Ala Ala Val Ala Ala Val Ala Ala Ala Asn Arg Ser  
 225 230 235 240

Ala Ser Ala Asp Gly Cys Ser Asp Arg Gly Ser Glu Arg Gly Thr Leu  
 245 250 255

Glu Arg Thr Asp Ser Arg Asp Asp Leu Leu Gln Leu Asp Tyr Ser Asn  
 260 265 270

Lys Asp Asn Asn Asn Ser Asn Ser Ser Ser Thr Gly Gly Asn Asn Asn  
 275 280 285

Asn Asn Asn Asn Asn Asn Asn Ser Ser Ser Asn Asn Asn Asn Ser  
 290 295 300

Ser Ser Asn Arg Glu Arg Asn Asn Ser Gly Glu Arg Glu Arg Glu Arg  
 305 310 315 320

Glu Arg Glu Arg Glu Arg Asp Arg Asp Arg Glu Leu Ser Thr Thr Pro  
 325 330 335

Val Glu Gln Leu Ser Ser Ser Lys Arg Arg Arg Lys Asn Ser Ser Ser  
 340 345 350

Asn Cys Asp Asn Ser Leu Ser Ser Ser His Gln Asp Arg His Tyr Pro  
 355 360 365

Gln Asp Ser Gln Ala Asn Phe Lys Ser Ser Pro Val Pro Lys Thr Gly  
370 375 380

Gly Ser Thr Ser Glu Ser Glu Asp Ala Gly Gly Arg His Asp Ser Pro  
385 390 395 400

Leu Ser Met Thr Thr Ser Val His Leu Gly Gly Gly Asn Val  
405 410 415

Gly Ala Ala Ser Ala Leu Ser Gly Leu Ser Gln Ser Leu Ser Ile Lys  
420 425 430

Gln Glu Leu Met Asp Ala Gln Gln Gln Gln His Arg Glu His His  
435 440 445

Val Ala Leu Pro Pro Asp Tyr Leu Pro Ser Ala Ala Leu Lys Leu His  
450 455 460

Ala Glu Asp Met Ser Thr Leu Leu Thr Gln His Ala Leu Gln Ala Ala  
465 470 475 480

Asp Ala Arg Asp Glu His Asn Asp Ala Lys Gln Leu Gln Leu Asp Gln  
485 490 495

Thr Asp Asn Ile Asp Gly Arg Val Lys Cys Phe Asn Ile Lys His Asp  
500 505 510

Arg His Pro Asp Arg Glu Leu Asp Arg Asn His Arg Glu His Asp Asp  
515 520 525

Asp Pro Gly Val Ile Glu Glu Val Val Val Asp His Val Arg Glu Met  
530 535 540

Glu Ala Gly Asn Glu His Asp Pro Glu Glu Met Lys Glu Ala Ala Tyr  
545 550 555 560

His Ala Thr Pro Pro Lys Tyr Arg Arg Ala Val Val Tyr Ala Pro Pro  
565 570 575

His Pro Asp Glu Glu Ala Ala Ser Gly Ser Gly Ser Asp Ile Tyr Val  
580 585 590

Asp Gly Gly Tyr Asn Cys Glu Tyr Lys Cys Lys Glu Leu Asn Met Gln  
595 600 605

Arg Asn Ile Arg Cys Ser Arg Gln Gln His Met Met Ser His Tyr Ser  
610 615 620

Pro His His Pro His His Arg Ser Leu Ile Asp Cys Pro Ala Glu Ala  
625 630 635 640

Ala Tyr Ser Pro Pro Val Ala Asn Asn Gln Ala Tyr Leu Ala Ser Asn  
645 650 655

Gly Ala Val Gln Gln Leu Asp Leu Ser Thr Tyr His Gly His Ala Asn  
660 665 670

His Gln Leu His Gln His Pro Pro Ser Ala Thr His Pro Ser His Ser  
675 680 685

Gln Ser Ser Pro His Tyr Pro Ser Ala Ser Gly Ala Gly Ala Gly Ala  
690 695 700

Gly Ser Val Ser Val Ser Ile Ala Gly Ser Ala Ser Gly Ser Ala Thr  
705 710 715 720

Ser Ala Pro Ala Ser Val Ala Thr Ser Ala Val Ser Pro Gln Pro Ser  
725 730 735

Ser Ser Ser Thr Gly Ser Thr Ser Ser Ala Ala Ala Val Ala Ala Ala  
740 745 750

Ala Ala Ala Ala Asn Arg Arg Asp His Asn Ile Asp Tyr Ser Thr  
755 760 765

Leu Phe Val Gln Leu Ser Gly Thr Leu Pro Thr Leu Tyr Arg Cys Val  
770 775 780

Ser Cys Asn Lys Ile Val Ser Asn Arg Trp His His Ala Asn Ile His  
785 790 795 800

Arg Pro Gln Ser His Glu Cys Pro Val Cys Gly Gln Lys Phe Thr Arg  
805 810 815

Arg Asp Asn Met Lys Ala His Cys Lys Ile Lys His Ala Asp Ile Lys  
820 825 830

Asp Arg Phe Phe Ser His Tyr Val His Met  
835 840

## IT IS CLAIMED:

1. A substantially isolated FRU polynucleotide.
- 5 2. The polynucleotide of claim 1, wherein the polynucleotide is selected from the group consisting of RNA, cDNA and genomic DNA.
3. The polynucleotide of claim 1, wherein the polynucleotide is derived from an insect that is a member of the phylum Arthropoda.
- 10 4. The polynucleotide of claim 3, wherein the polynucleotide is derived from an insect selected from the group consisting of medfly, fruit fly, tse-tse fly, sand fly, blowfly, flesh fly, face fly, housefly, screw worm-fly, stable fly, mosquito, and northern cattle grub.
- 15 5. The polynucleotide of claim 3, wherein the polynucleotide is derived from an insect that is a member of the order Diptera.
6. The polynucleotide of claim 5, wherein the polynucleotide is derived from a *Drosophila* polynucleotide.
- 20 7. The polynucleotide of claim 6, wherein the polynucleotide contains a sequence selected from the group consisting of SEQ ID NO:9 and SEQ ID NO:14.
8. A substantially isolated FRU polypeptide.
- 25 9. The polypeptide of claim 8, wherein the polypeptide is derived from an insect that is a member of the phylum Arthropoda.
10. The polypeptide of claim 9, wherein the polypeptide is derived from an insect selected from the group consisting of medfly, fruit fly, tse-tse fly, sand fly, blowfly, flesh fly, face fly, housefly, screw worm-fly, stable fly, mosquito, and northern cattle grub.
- 30 11. The polypeptide of claim 9, wherein the polypeptide is derived from an insect that is a member of the order Diptera.

12. The polypeptide of claim 11, wherein the polypeptide is derived from a *Drosophila* polypeptide.

13. The polypeptide of claim 12, wherein the polypeptide contains a sequence selected  
5 from the group consisting of SEQ ID NO:10 and SEQ ID NO:15.

14. A method of identifying a compound effective to alter the reproductive behavior of a target insect, comprising

10 treating an insect cell with a test compound, where said cell is obtained from the target insect and carries an expression vector containing FRU regulatory sequences operably linked to a reporter gene,

15 evaluating the level of expression of the reporter gene in the treated cell, and identifying the compound as effective if said compound significantly decreases the expression of the reporter gene in the treated cell relative to the expression of the reporter gene in untreated cells carrying said expression vector.

16. The method of claim 14, wherein the reporter gene encodes a protein selected from the group consisting of chloramphenicol acetyl-transferase (CAT),  $\beta$ -galactosidase ( $\beta$ -gal) and luciferase.

20

16. The method of claim 14, wherein the target insect is a *Drosophila* species, and the cells are selected from the group consisting of Schneider's Line 2 and *Drosophila* Kc cells.

17. The method of claim 14, wherein the regulatory sequences are from *Drosophila*.

25

18. The method of claim 14, wherein the target insect is a member of the phylum Arthropoda.

30

19. The method of claim 18, wherein the target insect is a member of the order Diptera.

20. The method of claim 18, wherein the target insect is selected from the group consisting of medfly, fruit fly, tse-tse fly, sand fly, blowfly, flesh fly, face fly, housefly, screw worm-fly, stable fly, mosquito, and northern cattle grub.

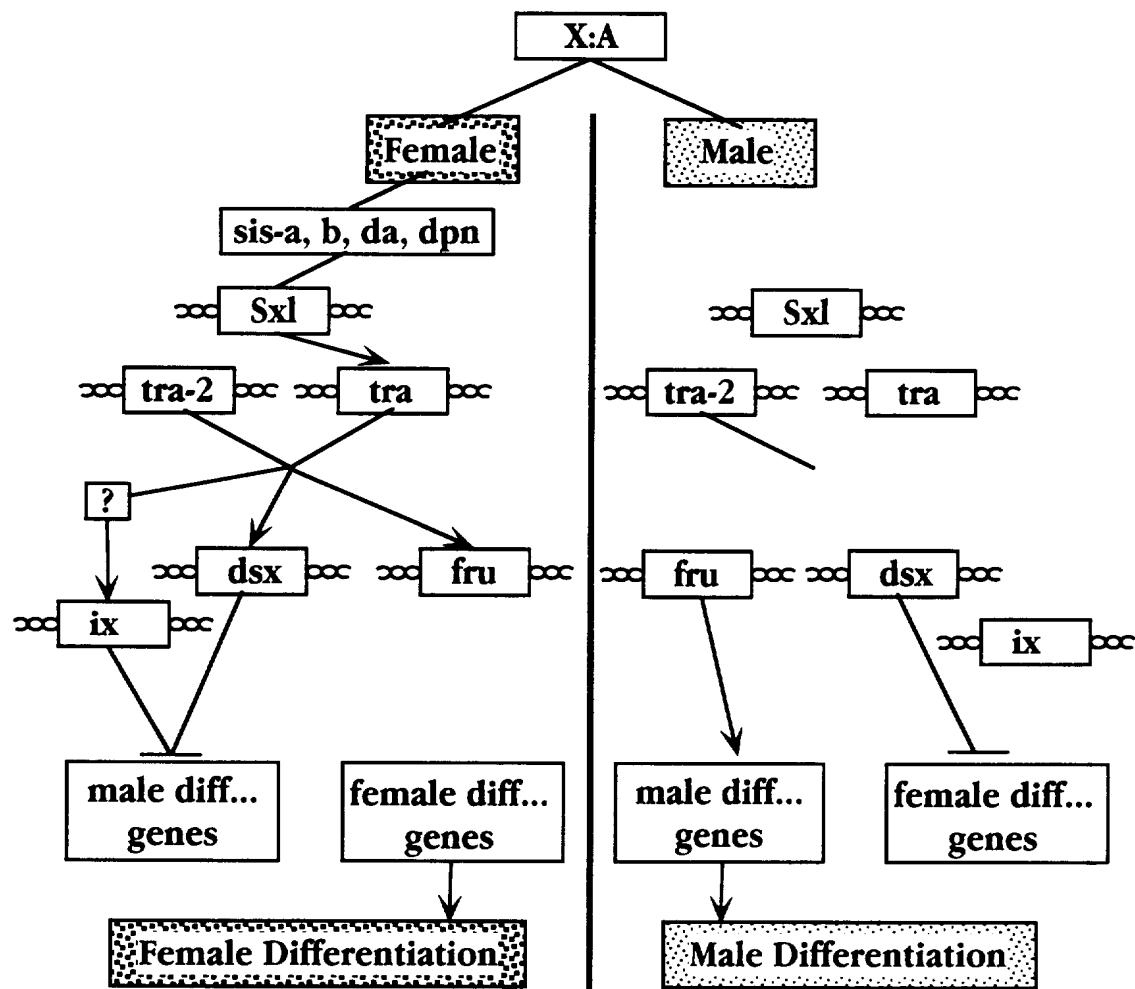
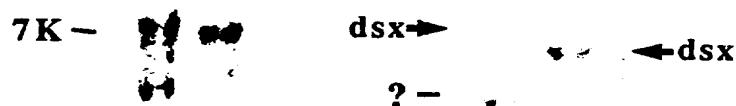
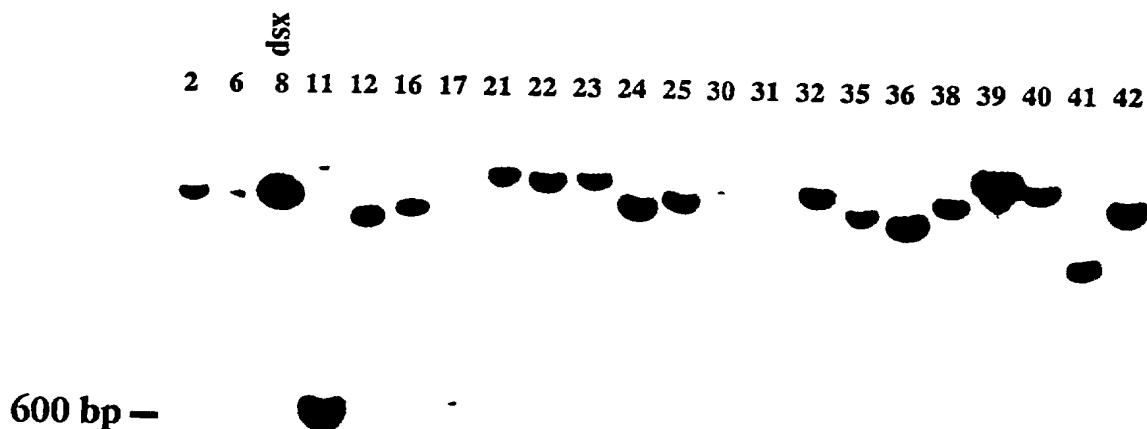


Fig. 1

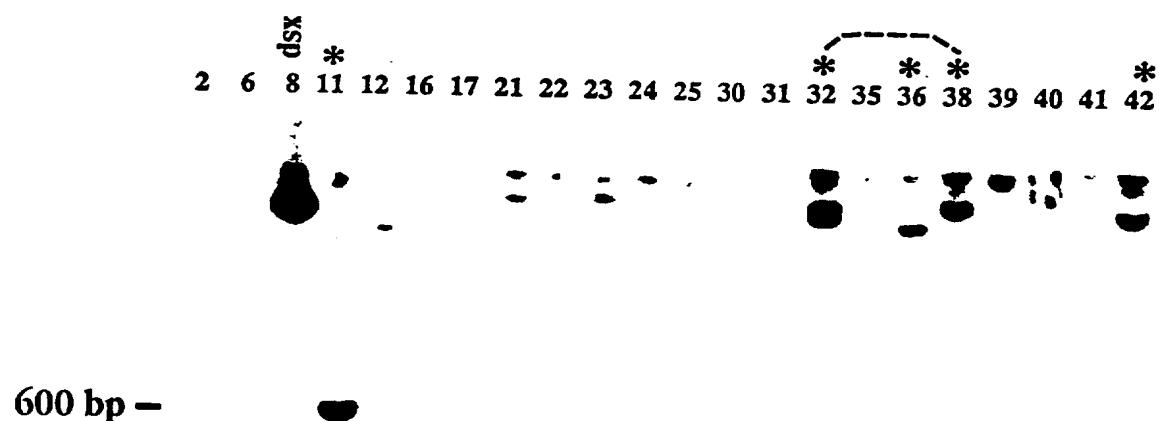
2/11

**47°C**BamHI  
EcoRI**51°C**BamHI  
EcoRI**2K-****600 -****- ?****Fig. 2A****Fig. 2B**

3/11



**Fig. 3A**



**Fig. 3B**

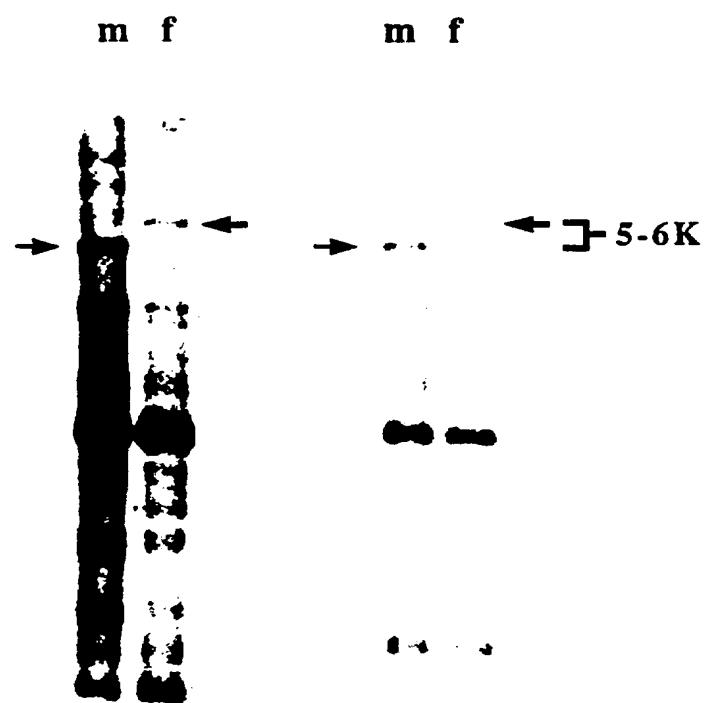
4/11

**EcoRI**

GAATTGAGGACGTGTGACGATGGAGCAACCCTCCCCCCCAGA  
TCGAAAGAGAATATCATCAATCAACATTCGGTCCCCGGAGGAG  
CTGCTCTTCAATCAACACTCAACCCGAACGGCCCTCAAAAGC  
CCGGCAACCTAAAGTTAGTCTTCATTAGCCTCTCTATCAATT  
AGTTAGTCAGCCAACGTTCTCTCTCTCATAATTCTAACCGA  
AAGTAAGCATAGAAAAGAACCAATACTCAATCAACATACCCAC  
AAAAAAAAACAAATCCCCACCAACTGGCGTCGGTAAGTGAAGAG  
CCATTAAATTATAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
NNNNNNNNNNNNNNNNNNNNNTGATGCCGATGATGCATGT  
GATAAGCAAGTGTGAACAATCCGTAGCAATCAGGCAGTAGGNN  
NCTTGAACAAATTAACTTAGCTGNATTTGCGCATGCCAAATG  
AAAAATAACAAACCGTAAATTCCAATGGTAACTAAACTAGCAA  
TACTAACTCTAGCCGATGGAACATGCAACCGAATT

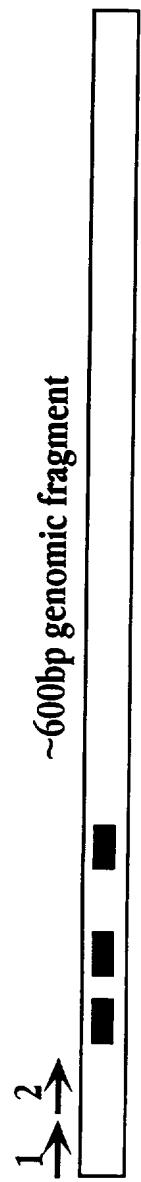
**EcoRI****Fig. 4**

5/11

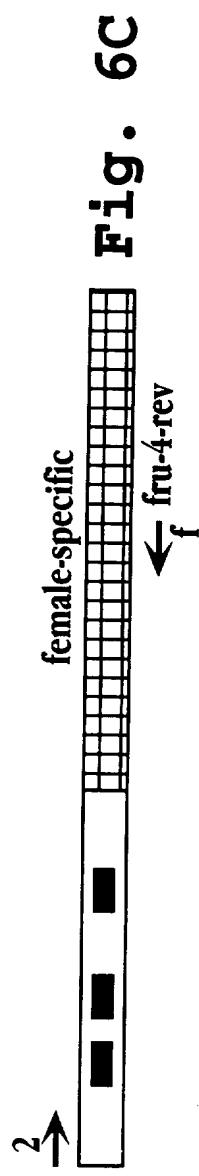
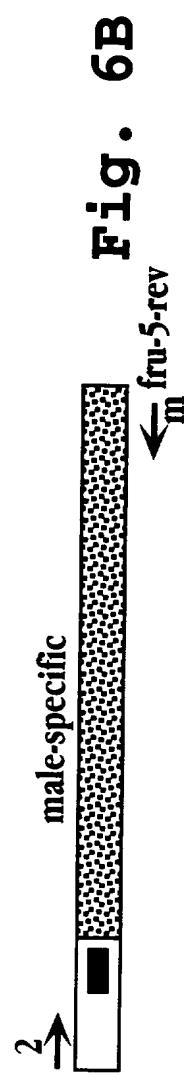


**Fig. 5A      Fig. 5B**

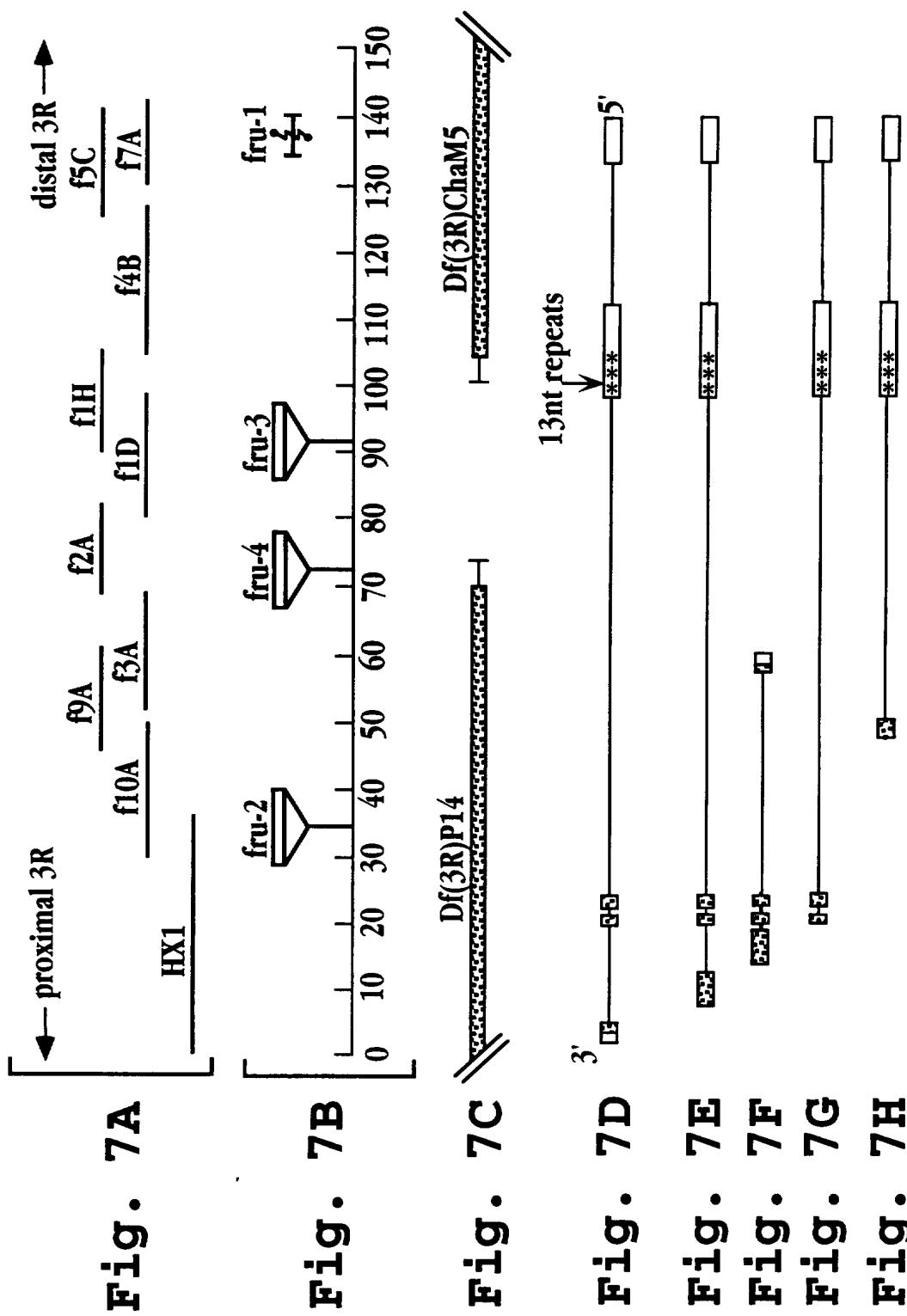
6/11

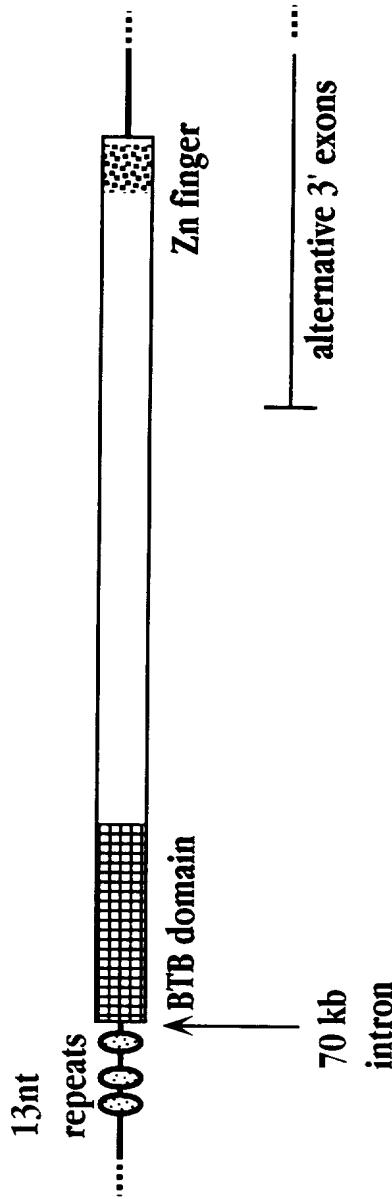


3' RACE products:



7/11





**Fig. 8**

9/11

gaattcggcacgaggattcacctatggcatatcagcaaacacatcaacgcacttctgttatgtctgcaatcaacc  
aaaatataaaaaaaaagaaaaaaaacccatcaactttaaagttttacgtttggtaaaggatttaaa  
atgccttaacttaacgcggccaaaggtaaacgttagattaaagttataggcaatcaatcgtaaaatatcgctttcg  
tttttaaaacttaccaatggactttgtatcccatcaattgcaatctaaagttagaaaaatagagagataagagatata  
atatcactaacaaagggtttcaggtataaaatgttaacttactacaatagaataacgtatttttttttttttttttttttt  
tagtatgtataaaggcaaaactgtcaagaaactgtcaaaacttacgttttttttttttttttttttttttttttttttttt  
tatcatatgttttaaggcagaggcatttacgacgacttgcgtttagactttaaacttatactatacacaatccatc  
gcaatccaaataacaataactaaggtagtt  
tattacttagataaaacatataatgtattatataatgttt  
gataaagaacacgactaaccatataacttactttagaaataacttatacgttttttttttttttttttttttttttt  
gatatcttatataatcggattt  
acacttttcaaggcaataaccataaaataatataaaggtaacgcggccaaagggttttttttttttttttttttttt  
cctaaacactacttagtgccttaaagtcttaggcataaccgcggccacccacggcggatgttttttttttttttt  
attatccaaggatacaactgttaaaggaaacccctatt  
atgttgcgttatcgttaaggcaatgttgcgttatcgttttttttttttttttttttttttttttttttttttt  
gtgtgacgtggacaacccttt  
tcttcaatcaacactcaacccggcaactggggcccttttttttttttttttttttttttttttttttttttttt  
caatt  
tcaacataccacaaaatcccaaaaacaaaatcccaaaaatcccaaaaatcccaaaaatcccaaaaatcccaaaaatcc  
CTGCTTGGCTGGAAACAATCATCCACAAATTGACCGGGCGTGCTAACCTCACTGGCTATGGACG

**Fig. 9A**

10/11

TCACGGCTCCCTGGAGGGCCAAACAGTCAAGGCTCACCGACCATCCTGTCAGGCCCTGCACTTGCAGTCCGACTT  
 TTCCCTACAGAACGCAATCCACATCCCATCATCTACTTGAAGATGTCAGATACTCAGAGATGGATCTGCTCGACTT  
 CATGTTACAGGGCGAGGTCAACGTTGGCCAGAGTTCGCTGGCCATGTTCTCAAGACGGCCGAGAGCCCTGCAAGGTGGT  
 GTCTCAACAGATAACACAATCTGAACACTACCGCTTCCGACTTCGACAAAGCTGGCGATTCGGGGCAGTTCGGGACCGGA  
 CGTGGCCGAGTAATTACACTGGGGCTTGGGGGGCTGG  
 GGGCTCCCGCTGGCAACGGGATCTGGGGGACGGACTGCGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG  
 CAGCAGG  
 ACCACGGCCACCGGG  
 ATGCAAGGATTCGGGGGAAAGGAAACGGGAAACGG  
 ACAAGGATAACAACATAAGCAACAGGACTAGTACCCGGGCAACAAATAAGGGAGGGAAACGGTGAAGGGAGGGGGGGGGGG  
 ACCAACAAACAACACCAGGAGCAATAGGGAGGG  
 GGGGAGACAGGGCACAGGGAGGCTGTCCACCCACGG  
 CCAACTGTGATAACTGGCTGGTCTGGAGCCACCCAGGGACAGGGCACTACGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG  
 CCCGGCCAAAACGGGGGGGAGCACATCGGAATCGGAGGG  
 AAGGGTTTCATCTGG  
 AGCAGGAGGTGATGGCACGG  
 GGGCTCTAAAGCTGCACGGGGAGGATAATGTCAACGGCTGGGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG  
 GCACAAACGAGCCAACAAACTGGGCTGG  
 CCCCTGTCGACCTCGTGGCTGG  
 CATCCCTCTGCACTATGG  
 TGCGGGCGAATAATTGGCATCACACCATGTTCAACGG  
 CAGCTGGCCATAACTGGTCACTCCACACGG  
 TGCGGGCTCCCTGGGGCAAGGAGGACTTCACTGGTGGGACCAATCGGTTGGGACCACTTGGTGGTCCATGGTGTCCCTA  
 CTGG

Fig. 9B

11/11

Fig. 9C

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US96/02331

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :536/23.5, 24.1, 24.31; 435/6, 70.3, 172.3; 424/9.2; 530/350; 436/501

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5, 24.1, 24.31; 435/6, 70.3, 172.3; 424/9.2; 530/350; 436/501

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, BIOSIS, DERWENT WORLD PATENT INDEX

search terms: fru, fruitless, gene cdna, locus

EMBL, GENBANK, EST/STS, GENSEQ DNA & PROTEIN DATABASES: SEQ ID NO.S 9, 10.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	RYNER et al., fruitless Lies at a Branch to the Sex-Differentiation Regulatory Hierarchy. In Program and Abstracts Volume, 35th Annual Drosophila Research Conference. Chicago, Illinois. 20-24 April 1994, page 32.	1-20
A	TAYLOR et al. Behavioral and Neurobiological Implications of Sex-Determining Factors in Drosophila. Developmental Genetics. 1994, Vol. 15, pages 275-296, especially 281, 282, and 293.	1-20

<input checked="" type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
-------------------------------------	--	--------------------------	--------------------------

• Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means		
*P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
23 MAY 1996	12 JUN 1996

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer CHARLES C. P. RORIES
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

**INTERNATIONAL SEARCH REPORT**International application No.  
PCT/US96/02331

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ZOLLMAN et al. The BTB Domain, Found Primarily in Zinc Finger Proteins, Defines an Evolutionarily Conserved Family That Includes Several Developmentally Regulated Genes in <i>Drosophila</i> . <i>Proceedings of the National Academy of Sciences</i> . October 1994, Vol. 91, pages 10717-10721, see entire document.	1-20
A	GAILEY et al. Behavior and Cytogenetics of <i>fruitless</i> in <i>Drosophila melanogaster</i> : Different Courtship Defects Caused by Separate, Closely Linked Lesions. <i>Genetics</i> . April 1989, Vol. 121, pages 773-785, see entire document.	1-20
A	GAILEY et al. Elements of the <i>fruitless</i> Locus Regulate Development of the Muscle of Lawrence, a Male-Specific Structure in the Abdomen of <i>Drosophila melanogaster</i> Adults. <i>Development</i> . 1991, Vol. L13, pages 879-890, see entire document.	1-20

**INTERNATIONAL SEARCH REPORT**

International application No. PCT/US96/02331
---

**A. CLASSIFICATION OF SUBJECT MATTER:**  
IPC (6):

C07H 21/04; C12N 15/12, 15/63, 15/85; C07K 14/435; C12Q 1/68; C12P 21/02; A61K 49/00; G01N 33/50, 33/68